SCIENTIFIC AMERICAN

NOVEMBER 1992
$3.95

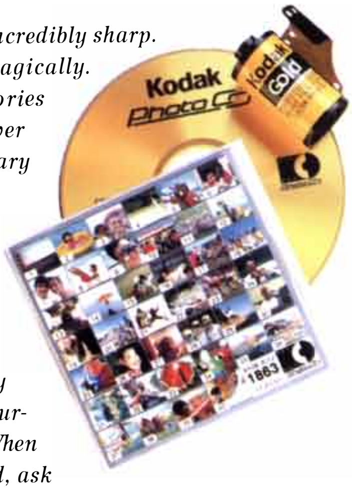
How Columbus's errors enlightened astronomy.
The puzzling big bang of animal evolution.
Software "glitches" that endanger public safety.

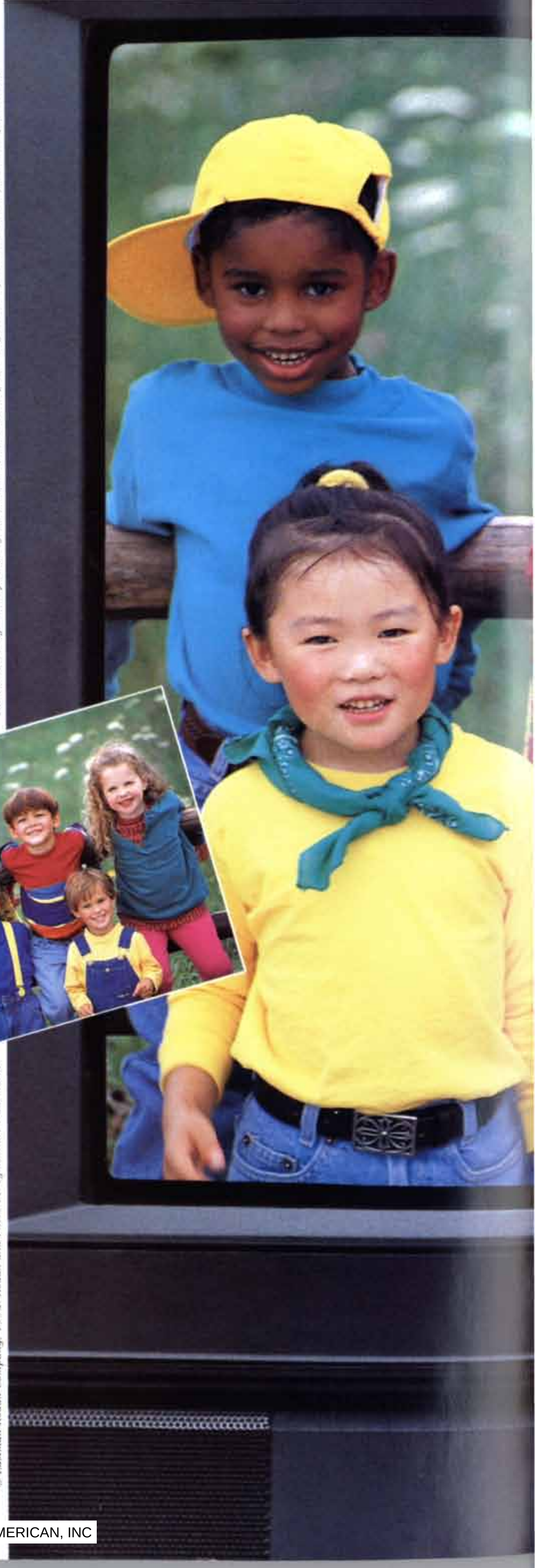
Tools for ants? *Not yet, but such gears may soon be components of useful machines measured in microns.*

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

Larger than life. Incredibly sharp. Glowing almost magically. The next generation of memories will be like nothing you've ever seen, thanks to a revolutionary new technology. It's called the KODAK Photo CD System. And it puts your favorite pictures into a powerful new showcase—your television. ¶ ***Amazingly Simple.*** *Keep taking pictures the way you always have, using your current 35mm camera and film. When you have your film developed, ask the photofinisher to transfer your photos onto a KODAK Photo CD Disc. Play the CD on a special KODAK Photo CD Player—and tell everybody to gather around.* ¶ ***Brilliant Images.*** *On your TV screen, you'll see high-resolution images, so vibrant you won't believe your eyes. Big, vivid pictures perfect for viewing with family and friends. Pictures you can actually change on-*

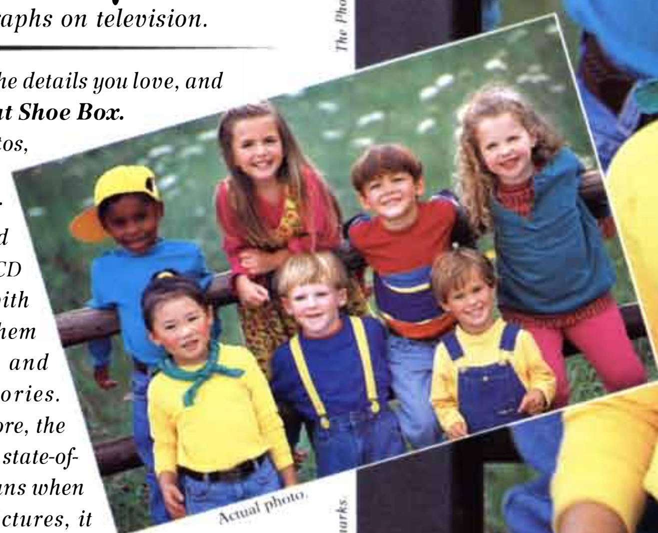




Introducing the KODAK Photo CD System.

It puts your photographs on television.

screen—resize them, zoom in on the details you love, and save the changes. ¶ ***Get Out That Shoe Box.*** *Now you can collect those old photos, slides and negatives gathering dust in your closet and have your photofinisher create customized Photo CD picture albums. Each CD holds up to 100 pictures, and with the remote control you can view them in any order. It's a convenient, and lasting, way to save your memories.* ¶ ***Sound and Vision.*** *What's more, the KODAK Photo CD Player is also a state-of-the-art audio CD player. That means when the Player isn't playing your pictures, it can play digital stereo sound. So whether you choose to watch your memories or listen to your favorite music, the KODAK Photo CD Player is ready to play both.* ¶ ***The Future of Memories.*** *This next generation of memories will be the biggest, brightest, most unforgettable ever. You've got to see it. Ask your retailer for a demonstration. For the retailer nearest you, call 1-800-945-7200.*

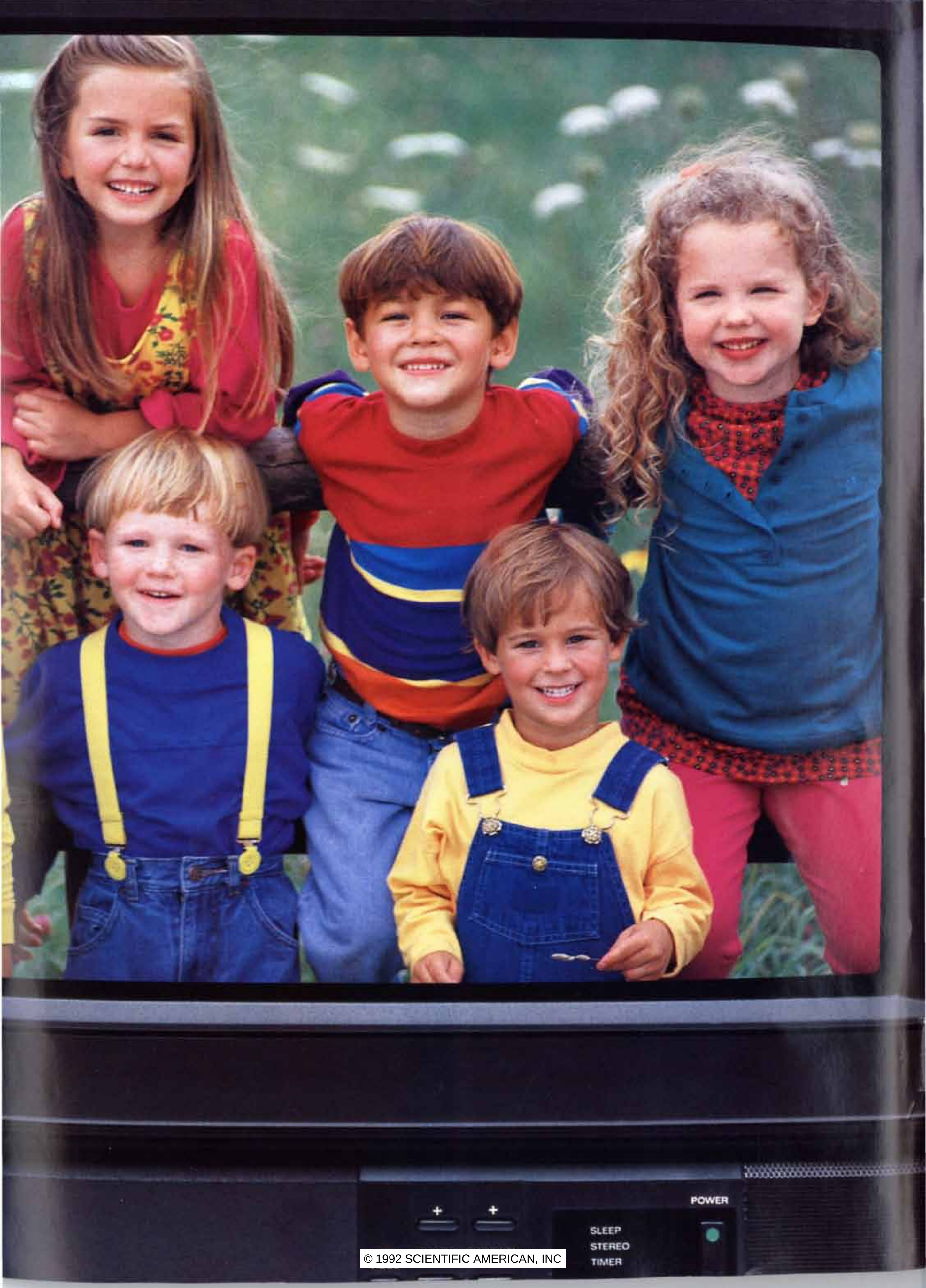


Actual photo.

Your Pictures Have Never Been So Powerful.™

Kodak

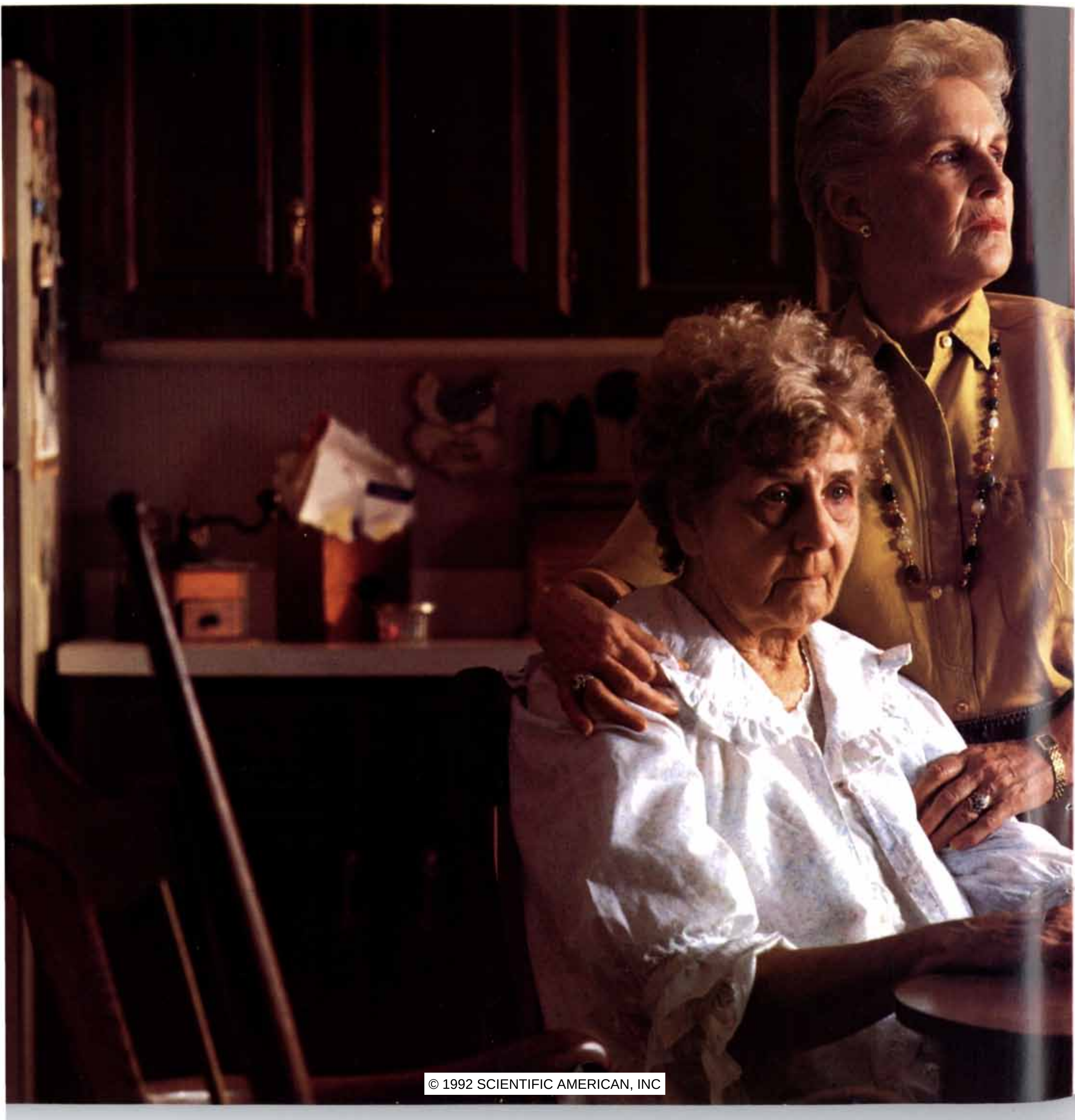
© 1992 SCIENTIFIC AMERICAN, INC



© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

Can we arrest t robs people o



© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

SCIENTIFIC AMERICAN

November 1992 Volume 267 Number 5

© 1992 SCIENTIFIC AMERICAN, INC

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 1992 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax: (212) 355-0408.

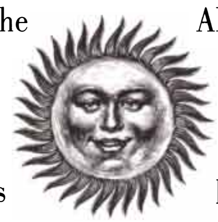
DEPARTMENTS

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

Plus *ça change, plus c'est la même chose*, the more things change, the more they remain the same. Lexus engineers are firm believers in this quip, at least when it comes to the way the automatic climate control of the Lexus ES300 sports sedan operates.

Friday *Saturday* *Sunday*



Four sensors, including a solar sensor that can measure ambient light from the sun's rays, are used to help gather information. This information is then relayed to the climate control microprocessor, which has at its disposal sixteen air vents in order to either cool you down or warm you up.

All this may sound highly technical (indeed it is), but it runs like a breeze. The temperature readout is an illuminated liquid crystal display; the controls are large and conveniently placed; the airflow warms or cools with remarkable quickness and efficiency.

The result? The weather may not be predictable, but the cabin of the Lexus ES300 surely is. Day in, day out, it's always the same old story. Which, when you think about it, is quite a pleasant surprise.

LEXUS

The Relentless Pursuit Of Perfection.

© 1992 SCIENTIFIC AMERICAN, INC

THE COVER image of an ant with a nickel gear for a mite machine wrapped over a leg was taken with a scanning electron microscope. The micrograph was made at the Karlsruhe Nuclear Research Center in Germany. The gear, 260 microns in diameter and 150 microns high, was made by a process called LIGA, the German acronym for a lithographic, electroplating and molding technique that is used to make microscopic mechanical components (see "Micron Machinations," by Gary Stix, page 106).

THE ILLUSTRATIONS

Cover image by the Institute for Microstructure Technology, Karlsruhe Nuclear Research Center. Computerized false coloring by Jason Küffer.

Page	Source
47	Jason Goltz; courtesy of Mount Sinai Medical Center, New York City
48	Will and Deni McIntyre/ Photo Researchers, Inc.
49-51	Jason Küffer
52	Johnny Johnson (*top*), Jason Küffer (*bottom*)
53	Connecticut General Life Insurance
54-55	Jim Riffle
56	Wendy L. Freedman
57	Kathleen Katimes/JSD
58-59	Jared Schneidman
60	Courtesy of Harvard University Archives
62-63	D. Aubert/Sygma
75	Nuclear Electric
76-77	Arthur J. Olson and David S. Goodsell
78-79	Richard E. Dickerson, University of California, Los Angeles (*left*), Teresa Larsen, U.C.L.A. (*center*), Arthur J. Olson and David S. Goodsell (*right*)
80	Arthur J. Olson (*top*), Victoria A. Roberts, Research Institute of Scripps Clinic (*bottom*)
81	Arthur J. Olson (*top left*), Elizabeth D. Getzoff and John A. Tainer, Scripps Clinic (*top center*), Michael Pique, Scripps Clinic, and David S. Goodsell (*top right*), Arthur J. Olson and David S. Goodsell (*bottom left and center*), Arthur J. Olson and I. Siara Mian, Scripps Clinic (*bottom right*)
85-88	Patricia J. Wynne
89	Johnny Johnson (*graph*), Patricia J. Wynne (*drawings*)
90	Patricia J. Wynne
91	John A. Endler, University of California, Santa Barbara
99	Michael Goodman (*left*), Arizona Historical Society Library (*center*), N. W. Tera/ Photo Researchers, Inc. (*right*)
101	*Jean de Dinteville and Georges de Selve* (*"The Ambassadors"*), by Hans Holbein, reproduced courtesy of the Trustees, National Gallery, London
102	*Nuremberg Chronicle;* courtesy of Owen Gingerich
103	Jana Brenning
104	*Fireball*, by Albrecht Dürer, by permission of the Syndics of the Fitzwilliam Museum, Cambridge
106-107	Electronics Design Center, Department of Electrical Engineering and Applied Physics, Case Western Reserve University
108-109	Gabor Kiss (*top*), Stephanie Rausser (*bottom left*), Berkeley Sensor and Actuator Center, University of California, Berkeley (*center and bottom right*)
110-111	Gabor Kiss; based on graphics from the Department of Electrical and Computer Engineering, University of Wisconsin–Madison (*top*), University of Wisconsin (*bottom left*), Neil McDonald (*bottom right*)
112	Robert Prochnow
113	Institute of Industrial Science, University of Tokyo (*left*), School of Electrical Engineering, Cornell University, and the National Nanofabrication Facility (*right*)

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, *Senior Writer;* Philip Morrison, *Book Editor;* Corey S. Powell; John Rennie; Philip E. Ross; Ricki L. Rusting; Russell Ruthen; Gary Stix; Paul Wallich; Philip M. Yam

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

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

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

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

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

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

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

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

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

PRESIDENT AND CHIEF EXECUTIVE OFFICER: John J. Hanley

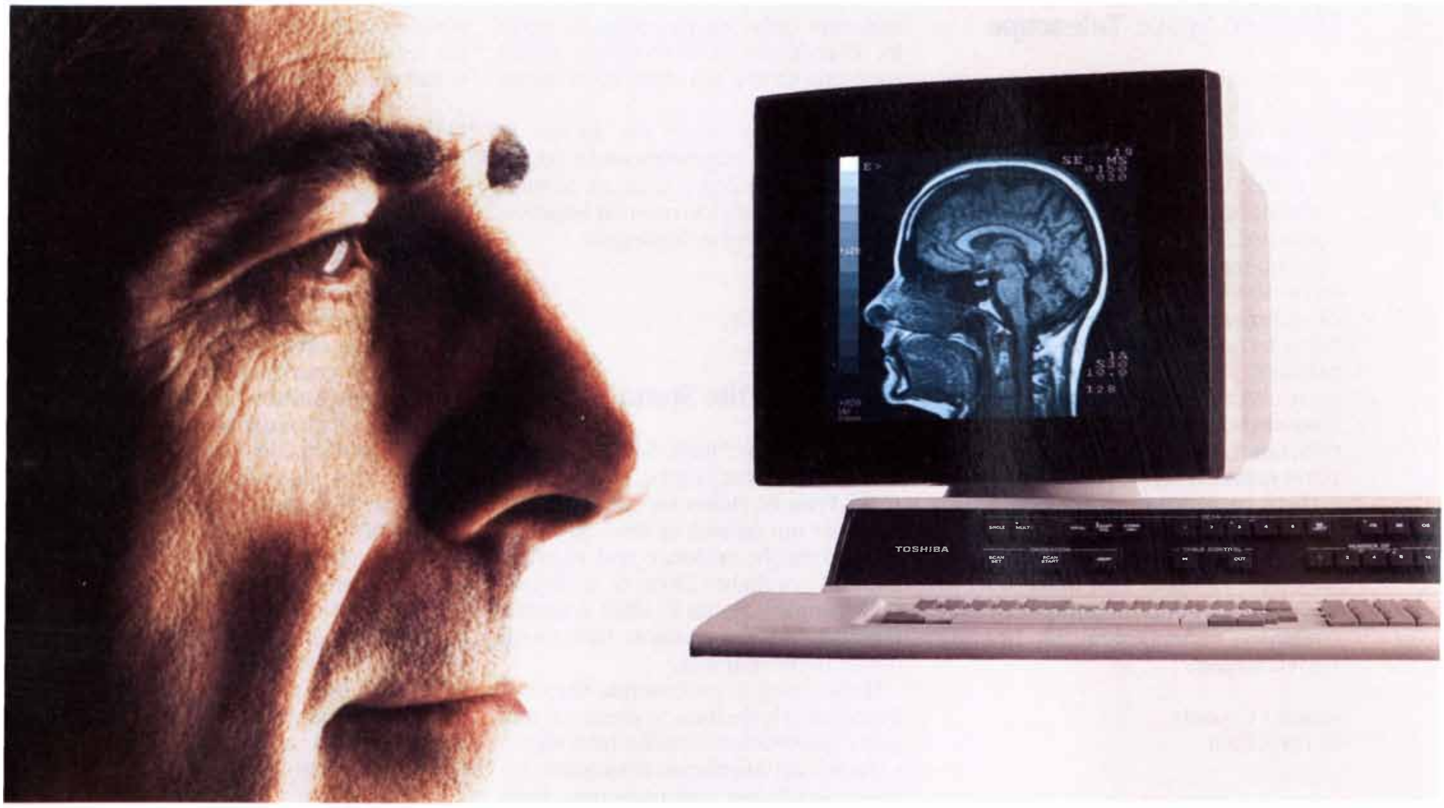
CORPORATE OFFICERS: *Executive Vice President and Chief Financial Officer,* R. Vincent Barger; *Senior Vice President,* Linda Chaput; *Vice Presidents:* Jonathan Piel, John J. Moeling, Jr.

CHAIRMAN OF THE BOARD: Dr. Pierre Gerckens

CHAIRMAN EMERITUS: Gerard Piel

© 1992 SCIENTIFIC AMERICAN, INC

NO ONE TRIES TO UNDERSTAND YOU MORE THAN WE DO.



It doesn't take the mind of a brain surgeon to understand people want to live longer, healthier lives. That's why Toshiba builds Magnetic Resonance Imagers that allow physicians to see deep within the human body without X-rays or exploratory surgery. Even abnormalities in cell structure, like cancer, can be discovered.

But, it did take the mind of a brain surgeon, as well as other medical experts at our research and development center in San Francisco, to help us design an even better MRI—our new MRT-150A. It's just one example of our pioneering role in super-conductive magnet technology and computer-image processing.

It's also an example of how much Toshiba cares about making the world a healthier place to live.

In Touch with Tomorrow

TOSHIBA

For more information write: Toshiba America Medical Systems, Inc., 2441 Michelle Drive, Tustin, CA 92681 or call: 1-800-421-1968.

© 1992 SCIENTIFIC AMERICAN, INC

LETTERS TO THE EDITORS

Hobbled Space Telescope

After reading Eric J. Chaisson's panegyric to the performance of the *Hubble Space Telescope* ["Early Results from the Hubble Space Telescope," SCIENTIFIC AMERICAN, June], I was struck by his enthusiasm for that flawed piece of hardware. The Hubble project was a financial, managerial and technical fiasco for NASA that resulted in the launch of an expensive orbiting telescope that has a technical capability far less than planned. The place to have found and corrected the fundamental design and construction errors was in the laboratory, not in orbit or in subsequent computer enhancements.

NASA spent many millions of dollars on *Hubble,* which ain't chopped liver. Unfortunately, *Hubble*'s performance ain't pâté. I wonder how much error or wishful thinking is incorporated in the computer enhancement of *Hubble*'s flawed imagery?

ROBERT C. GEISS
El Toro, Calif.

Will It Run on Time?

As a scientist, I am fascinated by maglev technology ["Air Trains," by Gary Stix; SCIENTIFIC AMERICAN, August]. As a commuter, I remain unimpressed. The goal of transportation planners should be to build high-speed conventional rail systems capable of moving commuters between cities 300 to 400 miles apart at 180 miles per hour—a reasonable compromise between the low-cost, low-speed automobile and the high-cost, high-speed airplane. If I could ride such a train for less than $120, that is the option I would choose. Maglev would have to be equally successful at balancing cost versus travel time, and I am not optimistic about that prospect.

MICHAEL TURBERG
Indianapolis, Ind.

The August issue, with its excellent article on maglev, also reported in "50 and 100 Years Ago" that on July 4, 1892, the Empire State Express made the 81-mile run from Rochester to Syracuse in 74 minutes. Amtrak now has a train of the same name, and its schedule calls for the train to make its 85-mile run in 78 minutes—within rounding errors, the same speed as its namesake.

Perhaps it is easier for society to make quantum improvements by adopting new technology, such as maglev, than by accreting incremental improvements in existing technologies.

EDWIN COHEN
Binghamton, N.Y.

Science on the Stand

In the essay "Junk Science in the Courtroom" [SCIENTIFIC AMERICAN, June], Peter W. Huber espouses that citizens are not capable of honestly evaluating scientific evidence and deciding the most probable cause of an injury. He essentially wants to stifle scientific thought and any opinion that might not be in the majority.

Huber uses as an example the drug Bendectin. He declines to point out that many epidemiologic studies have shown a statistically significant association between Bendectin and numerous birth defects. In fact, the Food and Drug Administration never concluded that Bendectin "did not cause birth defects." The truth is that every animal study ever done by anyone other than the manufacturer showed effects, from heart defects to hernias and missing limbs. The company's own studies also showed an effect on rabbits.

I suggest that it would be far better to have "junk science" in a courtroom than no science unless confirmed by governmental authorities.

BARRY J. NACE
Paulson, Nace, Norwind & Sellinger
Washington, D.C.

The relation of inept practices by obstetricians at the time of birth to later neurologic and motor problems, including cerebral palsy, is known and accepted in the field. Naturally, in particular cases, there will be contention among experts. The main source of "junk science" is from those such as Huber who make blanket pronouncements.

As a pediatrician and child psychiatrist with more than 20 years of field experience, I have seen at firsthand the effects denied by Huber. I fail to see what alternative Huber has to offer to the legal tradition of calling on expert witnesses in court cases.

MARCOSA J. SANTIAGO
Rumney, N.H.

Huber replies:

Individuals—particularly one such as Nace, who has spent many years litigating Bendectin cases—can believe many things, often with sincere conviction. But if courts are to resolve scientific controversies consistently and accurately, they must rely less on individual scientists, still less on individual lawyers and more on the published, peer-reviewed, consensus conclusions of scientific communities.

Readers interested in the science of Bendectin and cerebral palsy may refer to the FDA's published pronouncements on Bendectin, to the Institute of Medicine's *Medical Professional Liability and the Delivery of Obstetrical Care* (National Academy Press, 1989) and to the large body of scientific literature that those reports cite.

Jumbling the Genes

There was a mutation in "Genetic Algorithms," by John H. Holland [SCIENTIFIC AMERICAN, July], an otherwise fascinating introduction to computer programs designed to evolve. Chromosomal crossing-over, which leads to the recombination of genetic material, does not occur when sperm and ova fuse, as stated in the article. Rather crossing-over occurs during meiosis, the process that produces ova or sperm. This system allows a much greater degree of genetic recombination, and hence diversity, than would crossing-over at the time of fertilization.

MARY L. MARAZITA
Department of Human Genetics
Medical College of Virginia
Virginia Commonwealth University

All letters to the editor may be edited for length and clarity. Because of the volume of mail, letters cannot be acknowledged individually. Unsolicited manuscripts must be accompanied by a stamped, self-addressed envelope.

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

Over a million hav
and it's not eve

According to the polls, more than half a million people who used to use WordPerfect® have moved to Microsoft® Word for Windows.™

A million more Lotus® 1-2-3® users have moved to Microsoft Excel for Windows.

Like people all across the country, they were looking for a change, something that would make life a little easier.

That something is Microsoft.

© 1992 SCIENTIFIC AMERICAN, INC

Daniel Schorr, *National Public Radio,*
March 23, 1992

voted for change,
election day.

We make it possible to do everyday tasks with less effort. Writing a report, for example. Or developing a budget.

In fact, with Microsoft Excel and Word for Windows, just about anything is easier, short of running the country.

If you'd like to know more, call (800) 426-9400, Dept. HZ2. And keep in mind that change is about possibilities.

Microsoft

Making it easier

SA. Microsoft is a registered trademark and Windows is a trademark of Microsoft Corporation. Lotus and 1-2-3 are registered trademarks of Lotus Development Corporation. WordPerfect is a registered trademark of WordPerfect Corporation.

© 1992 SCIENTIFIC AMERICAN, INC

50 AND 100 YEARS AGO

NOVEMBER 1942

"Each week in *The Journal of the American Medical Association* there is a long series of obituary notices of deceased doctors. Each notice lists the dead doctor's life attainments. Naturally, some of these notices are long, some of medium length, others short, and it happens that, for reasons of appearance, the printer arranges the notices in order of length—longer ones preceding shorter ones. It occurred to a Brooklyn doctor that this weekly list, thus arranged, might provide an opportunity to determine 'what price success' in medicine. So he analyzed 30 such weekly lists and found that the average age of death of the first ten doctors on them was 64.6 years, while the last ten doctors—they who had served faithfully but not gained prominence—lived on to 69.3 years. Thus the price of success proves to be about five years of a doctor's life."

"The Germans were the first to grasp the tremendous possibilities of the theory that greater mobility in both offensive and defensive warfare could be secured by transporting troops and weapons by airplane. Their Junkers 52, a relatively slow and plodding but roomy and reliable type of plane, has been carrying supplies, weapons, infantry, and parachutists to Norway, to Libya, to the Russian front. Our own Air Corps, or let us say rather our Army, was a little slow in realizing the potentialities of air troop transport. Today, however, thanks in part to the conversion of some splendid transport airplanes, we are as well equipped as any; an endless stream of cargo ships is flying to China to replace the Burma road, to Northern Africa, and to other fighting fronts."

Alexander Bell, calling long-distance

SCIENTIFIC AMERICAN

NOVEMBER 1892

"As is well known, it is the common belief that the hairs of mammals, the feathers of birds, and the scales of reptiles are all epidermal structures of a fundamentally identical character, but after an elaborate study of the growth and development of these several protective coverings, Dr. F. Maurer, of Heidelberg, now arrives at the conclusion that they are homologous with the sensory points in the skin of the amphibia, or, at least, that they are outgrowths from these points as bases. Dr. Maurer thus concludes that his researches confirm that the mammalia are derived directly from the amphibia, and have not had any reptilian ancestors."

"An unfinished obelisk in the quarries at Syene shows the mode in which the ancients separated these monoliths from the native rock. In a sharply cut groove marking the boundary of the stone are holes evidently designed for wooden wedges. After these had been firmly driven into the holes, the groove was filled with water. The wedges swelled and cracked the granite throughout the length of the groove. The block was pushed forward upon rollers from the quarries to the edge of the Nile, where it was surrounded by a large timber raft. It lay by the river side until the next inundation floated the raft and conveyed the obelisk to the city where it was to be set up. There, by means of rollers, levers, and ropes, the obelisk was gradually hoisted into an upright position."

"In the opening of the telephone line between New York and Chicago as given in our last number, Prof. Bell was photographed by flashlight while talking with Mr. William H. Hubbard, at Chicago, a distance of nearly 1,000 miles. Our illustration, taken from the *Electrical Review,* is reproduced from the flashlight photograph and is interesting historically as showing the advances made in both sciences, telephony and photography."

© 1992 SCIENTIFIC AMERICAN, INC

THE ONLY BRAKE PEDAL THAT MAKES A SPORTS CAR GO FASTER.

The brake pedal in the new Mazda RX-7 may be the lightest in the world. We made it out of aluminum, then drilled holes in it. Why go to such extremes? We know what a handful of enthusiasts have always known. In a pure sports car every ounce counts. That's why we started with a sequential twin turbo rotary engine that's 200 pounds lighter than a comparable piston engine. Which gave us a head start on a weight reduction program nothing short of obsessive (we even trimmed the spark plug leads). The result is something that you may never see again. 2789 lbs. 255 horsepower. 0-60 in 4.9 seconds. And exclusive rights to the label pure sports car.

© 1992 SCIENTIFIC AMERICAN, INC

Declassified

Russian geophysicists seek new ways of making a living

Not so long ago a Russian spy plane bristling with detectors flying into the heart of the U.S. air defense network would have sent fighter jets scrambling. But this past September 12, radar operators shrugged when a massive Ilyushin 76-MD "flying laboratory" touched down at Denver's Stapleton airport, just a few minutes' flying time from, among other places, the North American Aerospace Defense headquarters in Cheyenne Mountain and the spy-satellite station at Buckley Air National Guard Base.

In case any further proof was needed that the cold war is over, the 32 scientists and 14 crew members, many of them former employees of the Soviet military complex, then threw open the cargo doors, brought out the vodka and invited all comers to crawl over the huge craft and inspect its gear. Cameras welcome.

MILITARY AND CIVILIAN U.S. scientists examine Russian Ilyushin 76-MD "flying laboratory" at Denver's Stapleton airport. Photo: Harry R. Olsson.

The landmark visit was the culmination of a seven-month effort by Warren T. Dewhurst, chief geophysicist of the coastal and geodetic survey of the National Oceanic and Atmospheric Administration (NOAA). Dewhurst has been forging links with his colleagues in the former evil empire for over a year and first saw the flying laboratory this past February at the Gromov Flight Testing Institute near Moscow. It was then, Dewhurst says, that he conceived of bringing it to the U.S. as an advertisement for the Geophysical Technology Transfer Initiative that he was planning with Norman Harthill, a geophysicist at the Colorado School of Mines in Golden, and Serguej N. Domaratskij, a researcher at the St. Petersburg Institute for Terrestrial Magnetism, Ionosphere and Radio Wave Propagation. Tomas G. Musiniantz of the Institute for Precise Instrumentation in Moscow suggested enlisting Russian scientists for the trip to Colorado.

Dewhurst, Harthill and Domaratskij are convinced that U.S. and Russian geophysicists have a lot to teach one another because they have worked separately for decades. The principal obstacle to cooperation has been the sensing technologies they use, which until recently were classified. Although surveys of minute variations in the earth's magnetic and gravitational fields can be used to locate minerals, oil and gas, magnetic-field disturbances over the ocean can also reveal the presence of submarines. Gravity-field maps can be used for accurate targeting and can pinpoint underground caverns that conceal missiles.

The magnetometers on the specially modified Ilyushin are of military design, according to Harthill, who says the craft is "basically a Soviet military antisubmarine warfare plane." The four-jet engine aircraft, which has a range of 8,200 kilometers and can accommodate 40 tons of cargo, carries synthetic-aperture imaging radar as well as gravimeters. NOAA's fleet of ships and small aircraft carry no such instruments. "Our ships go out and measure depth, and that's it," Dewhurst says ruefully.

Dewhurst says he does not embarrass his Russian counterparts by asking them about military missions they may have flown. "If this collaboration is mutually acceptable, we should allow it to happen," he says. The U.S. military was not so eager, however. Securing landing rights for the Ilyushin in Denver took a major effort, and Dewhurst also ran into "remnants of the cold war" in the State Department. The day before the airplane left Moscow, many of the crew and scientists still had not received visas to enter the U.S.

Once on the ground at Stapleton, the Russian scientists spent the next four days at the Colorado School of Mines making sales pitches to U.S. geophysicists—and to sharp-eyed military types from places like the Naval Surface Warfare Center and the Defense Mapping Agency. With military support drying up, the Russians were unabashed about making their appeals. "We have the experience, we have the means, we have the desire," said Musiniantz, winding up an overview of the flying laboratory. "What we are a little short of is the money."

Among the systems the Russians described was a powerful LIDAR (a laser-based radar that can be used to see below the sea surface) that employs a 300-kilowatt copper-vapor laser. According to Viktor I. Feigels of the Institute of Fine Mechanics and Optics in St. Petersburg, the system can provide useful information on depths down to 20 or 30 meters. Jon Davis of the Naval Air Warfare Center in Warminster, Pa., says the U.S. Navy might cooperate with the Rus-

© 1992 SCIENTIFIC AMERICAN, INC

From Totally Elegant

To Extra Crude

We don't make a lot of the products you buy. We make a lot of the products you buy better.

From the most beautiful clothing on earth, to the riches beneath, BASF is helping to make them happen.

Our textile products make fabrics softer and more desirable. They make colors more colorful while they improve wear.

BASF energy products help support crude oil production. They prevent corrosion and recover more energy by removing water from oil.

For a free booklet describing what BASF is into these days, call 1-800-669-BASF. We help make the products you buy better in some beautiful ways.

BASF

© 1992 SCIENTIFIC AMERICAN, INC

sians on LIDAR work "because I think they are a little better than we are in this area."

Many of the U.S. observers agreed that the Russians' relatively unsophisticated electronic equipment had forced them to solve problems in ingenious ways. They described many techniques for improving airborne gravity surveys, for example. "They have outstanding theoreticians," says Richard J. Wold of GWR Instruments in San Diego, which manufactures gravimeters. Some observers were disappointed. Although the visitors described on-board computing for processing synthetic-aperture radar data, "they didn't show any really nice synthetic-aperture images," says Andrew Ochadlick of the Naval Air Warfare Center. But he added, "I'm going to learn to speak Russian."

The Golden conference was the first formal event under the rubric of the Geophysical Transfer Initiative. The next will be a conference in 1993 in St. Petersburg, where U.S. geophysicists will describe their research. In addition, Dewhurst and Domaratskij have created a permanent foundation in St. Petersburg dedicated to joint geophysics projects.

Although the foundation is brandnew, Dewhurst already has one project for it in mind. He is trying to drum up support for acquiring (he is sketchy on financial details) a Russian "flying boat" designated the Beriev A-40. Two prototypes already exist. The aircraft would be equipped with gravimetric, LIDAR and other instruments and would be used for surveys in polar regions—or anywhere else. The airplane's long range and amphibious capability make it the ideal choice, Dewhurst and Domaratskij believe, because it could survey from the air and then land at sea to make "ground truth" measurements. The researchers already have one survey mapped out: the recent no-man's-land of the Bering Strait, where, they agree, there is a tunnel just waiting to be built. The stretch is a mere 60 miles—with an island in the middle—and the water is shallow, Dewhurst says.

In the meantime, Dewhurst would be delighted to see the Ilyushin conducting geophysical surveys in the U.S. Its onboard LIDARs make it ideal for air- and water-quality monitoring, Harthill points out. And to anyone who finds incredible the idea of the U.S. Air Force's allowing the multipurpose survey airplane to fly four-kilometer-spaced parallel lines over Kansas with its sensors switched on, Dewhurst has a ready reply. "A month ago the military wasn't keen about it landing in Colorado," he says, gesturing to the airplane on the tarmac. "But here it is." —*Tim Beardsley*

Paradise Lost?

Microbes mount a comeback as drug resistance spreads

Only a few years after penicillin moved into widespread use during the 1940s came the first reports that some bacteria had grown resistant. And as more powerful antibiotics such as streptomycin, tetracycline and chloramphenicol were developed, bacteria evolved resistance to them, too. Today bacteria and fungi that are resistant to once effective drugs are causing deaths and driving up the medical costs all over the world.

Drug resistance was mostly ignored in the U.S. until recently because physicians believed they had access to all the antibiotics they might need, says Stuart B. Levy, a researcher at Tufts University. They were wrong. Drug resistance has been found in virtually every type of microbe that has been fought with antibiotics. That covers everything from food-borne pathogens such as *Salmonella* to sexually transmitted organisms such as *Neisseria gonorrhoeae.* Surgical patients are now dying in U.S. hospitals from wound infections caused by enterococcal bacteria resistant to several different drugs. "These are people who should probably not be dying," says David Shlaes, a physician at the Veterans Administration Medical Center in Cleveland.

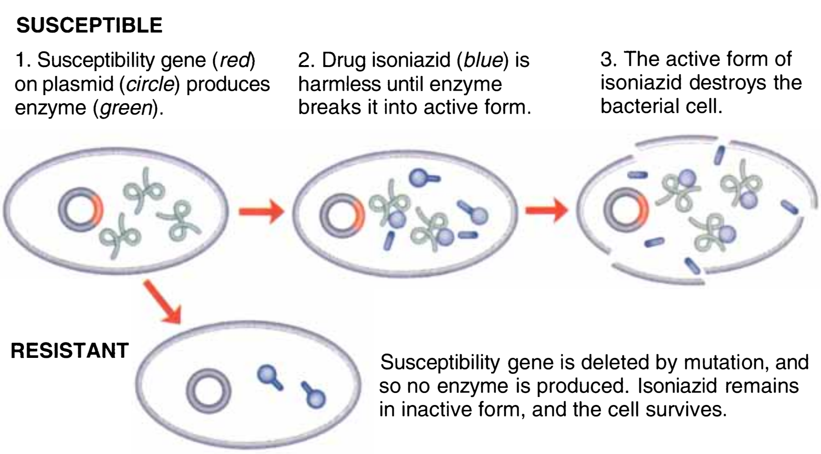
Although an infection that is resistant to one drug can often be cured by switching to a different (and usually more expensive) one, increasing numbers of pathogens are resistant to several drugs. During the 1980s, outbreaks of multidrug-resistant dysentery, cholera and pneumonia, to name just a few examples, were recorded around the world. And in the past two years the U.S. medical establishment has been shaken by the emergence of multidrug-resistant *Mycobacterium tuberculosis,* a scourge that many physicians assumed had been quashed in developed countries. During 1991, tuberculosis resistance to one or more drugs was reported in 36 states.

Resistance of TB to drugs represents an alarming threat because, unlike many other serious infections, "the principal risk behavior for acquiring TB infection is breathing," state Barry R. Bloom and Christopher J. L. Murray in a recent issue of *Science.* Bloom, an investigator at the Albert Einstein College of Medicine in the Bronx, and Murray, who works at the Harvard School of Public Health, point out that TB is the leading cause of death from infectious disease worldwide, with eight million new cases and 2.9 million deaths every year.

In the U.S., the number of TB cases has been increasing since 1985, with more than 26,000 reported in 1991. And in New York City, one third of all cases tested in 1991 were resistant to one or more drugs. Uncomplicated TB can be cured with a six-month course of antibiotics, but the outlook for multidrug-resistant cases is bleak. Those resistant to two or more major antibiotics have a fatality rate of around 50 percent; patients infected with the human immunodeficiency virus (HIV), the causative agent of AIDS, succumb in only a few weeks. According to Bloom and Murray, attempts to treat people who are HIV positive may have permitted the emergence of *M. tuberculosis* resistant to virtually all anti-TB drugs. The disease is also spreading rapidly in people not infected with HIV.

Many drug-resistant bacteria have

How Tuberculosis May Develop Drug Resistance



© 1992 SCIENTIFIC AMERICAN, INC

A special offer for Scientific American readers...

How can our hands strike strong blows, and yet be capable of delicate caresses?

What if our arms had extra joints or different proportions—would they be more useful?

What goes awry when someone develops arthritis or sprains an ankle?

What makes the complex mechanisms of human movement so unique that no robot can duplicate them exactly?

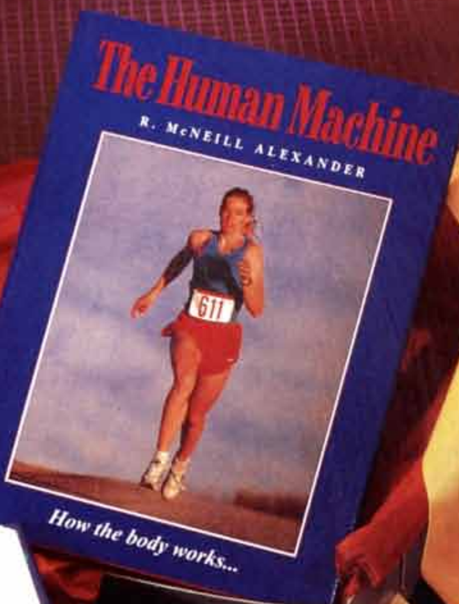
To answer these questions, R. McNeill Alexander examines the mechanical complexities of human motion as an engineer might. From breathing to drinking and reaching to swimming, Alexander devotes each chapter to a single action, raising and answering questions as he goes.

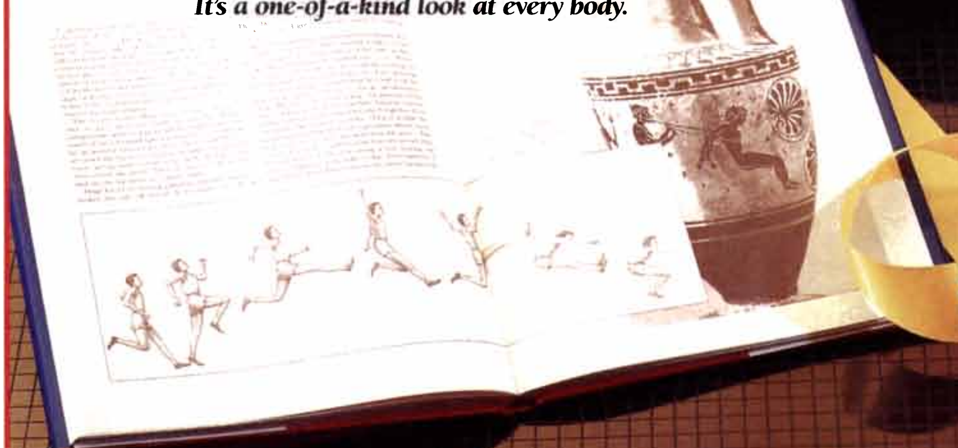
The variety of human movements covered in THE HUMAN MACHINE will captivate any reader. Sixty-three black & white and color photographs and 129 detailed illustrations highlight nearly every page, enhancing his explanations.

Well-known to *Scientific American* readers, Alexander is the editor of SA Library's EXPLORING BIOMECHANICS: ANIMALS IN MOTION and author of "How Dinosaurs Ran," a 1991 *Scientific American* cover story. This is his latest work.

THE HUMAN MACHINE makes the perfect addition to your collection, or an ideal gift for friends, family, and colleagues.

It's a one-of-a-kind look at every body.





CALL (800) 944-UNIV OR FAX (800) 944-1844
CREDIT CARDS ACCEPTED

"Alexander's broad method is the comparative engineering of large organisms, seen in the light of the physical necessities that link structure to function and of the style of design by evolution. The reader's pleasure and instruction spring from the directness of the work."

—SCIENTIFIC AMERICAN

"It is not easy to write simply, clearly and concisely. R. McNeill Alexander is a master of the art."

—NATURE

A special offer to SCIENTIFIC AMERICAN readers... You pay just $30.00 with FREE postage and handling. (The retail price of THE HUMAN MACHINE is $34.95.) Satisfaction is guaranteed or your money back!

Whether a scientist or science buff, you will thoroughly enjoy this engrossing new look at the body.

COLUMBIA UNIVERSITY PRESS

DEPT. T23 • 136 SOUTH BROADWAY, IRVINGTON, NY 10533

© 1992 SCIENTIFIC AMERICAN, INC

genes that enable them to produce enzymes that either destroy particular antibiotics or pump them out of the bacterial cells. Because bacteria pass genes not only between members of the same species but also between different species, genes that confer resistance to a widely used drug spread rapidly.

Until this summer, the mechanism of drug resistance to TB was unknown. Ying Zhang of the Hammersmith Hospital in London and his colleagues discovered the basis for resistance to a TB drug used frequently in treatment, isoniazid. Surprisingly, resistance in this instance seems to result from the loss of a gene, one on a plasmid, or loop of DNA. The gene probably allows the TB bacterium to process isoniazid into an active form. That finding suggests some promising avenues for research. But developing a new drug usually takes at least seven years, Shlaes points out.

Yet drug companies have been scaling back research on antibiotics and other antimicrobials and turning instead to anticancer and antiviral drugs. So when drug-resistant TB first came to national attention in 1991, there was no U.S. source for two TB drugs that were formerly used to treat the disease, streptomycin and *para*-amino salicylate sodium.

Thomas L. Copmann, head of biotechnology of the Pharmaceutical Manufacturers Association, places the blame on the "dismal" amount being spent on drug resistance by the National Institutes of Health. According to John R. La Montagne of the National Institute of Allergy and Infectious Diseases, the amount that institute spent on research related to drug resistance in 1991 was $9.7 million, up from $6.7 million in 1988. But Shlaes says much of that total is only peripherally related to the practical problem. And although the NIH has sponsored two workshops on drug resistance since 1989, the agency declined to support a major new research effort.

"If people are going to take it seriously, the NIH and the Centers for Disease Control will have to rebudget," Bloom says. He charges that the CDC knew of some cases of drug-resistant TB more than two years ago and considered—but rejected—a program to eliminate it. "If that had been done, we would have the infrastructure we need now," he says. "The tragedy is you have to have skeletons on the front page of the newspapers before people can be persuaded something must be done."

Physicians are setting up informal networks for monitoring antibiotic resistance. Levy is seeking NIH funding for a voluntary international data base. Thomas F. O'Brien and John M. Stelling of Harvard Medical School have developed a data management program that helps microbiology laboratories track drug-resistant organisms. That effort, which includes posting reports on an international network, is supported by the World Health Organization.

WHO is now helping several countries set up TB control programs. In the U.S., the CDC is starting to monitor drug resistance in different organisms, including *M. tuberculosis.* Dixie E. Snider of the CDC says the agency wants eventually to obtain data on drug susceptibility for every case of TB in the U.S., but that monitoring system is not yet running.

In the meantime, physicians could minimize the spread of drug resistance by not using antibiotics that are more powerful than necessary, says George Jacoby, an infectious disease specialist at Massachusetts General Hospital. About half of all antibiotic use in the U.S. is inappropriate, asserts Calvin M. Kunin, who chairs a committee on antibiotics for the Infectious Disease Society of America. Many prescribed antibiotics, he says, are either more powerful than necessary, used for longer than necessary or not needed at all. "It's a cacophony," Kunin declares. "We have a long way to go." *—Tim Beardsley*

Bringing Science to the Bottom Line

The enthusiasm in Washington for steering research toward the bottom line has reached the National Science Foundation (NSF), the traditional mainstay of science unsullied by commercialism. Walter E. Massey, the foundation's director, appointed a blue-ribbon commission in September to "examine ways for NSF to accept an enhanced role in fostering connections between research and technology." The group's industrial emphasis is underscored by the choice of Robert Galvin, former chief executive officer of Motorola, to serve as co-chairman with William H. Danforth, chancellor of Washington University.

Some advocates of basic research are unwilling to give up without a fight. The American Physical Society is "deeply concerned" that the NSF may be wavering in its commitment to basic science, according to spokesman Robert L. Park. Physicists plan to make their objections loud and clear.

Massey has been pushing for changes at the NSF for more than a year. He cites the end of the cold war, the increasingly international character of science and the downturn in U.S. corporate research as reasons why the agency, which has a budget of almost $2 billion, should reexamine its operation. Massey says he favors giving the agency "an expanded portfolio of programs that would be integrated with ongoing activities and closely aligned with industry and other government agencies."

When Massey presented his ideas this past June to the National Science Board, the NSF's governing body, the board declined to approve that kind of mission shift without first getting it blessed by an outside group. The commission is being asked to finish its report by November, so that whichever administration is in the White House next year will have a blueprint for action. The group's recommendations are likely to find receptive ears on the National Science Board and in Congress. Both bodies have called for more applied research.

The NSF's navel contemplation comes hard on the heels of a similar effort to develop a strategic plan for the National Institutes of Health, the principal avenue of federal support for basic biomedical research. The NIH, with a budget of some $9 billion, has traditionally supported a mixture of pure basic research and studies aimed at specific diseases. Development of the NIH plan, which director Bernadine P. Healy initiated more than a year ago, engendered apprehension and outright hostility from some researchers. NIH sources say Healy's plan has also run into trouble with her political taskmasters at the Department of Health and Human Services, who view it as an attempt to justify further increases in the NIH's budget.

According to some NIH prognosticators, the long-gestating strategic plan will be published, with suitable bows to academic freedom as well as the need for commercial development, and then ignored. The commission on the NSF will, presumably, be aiming to ensure that fate does not befall Massey's plan. *—Tim Beardsley*

© 1992 SCIENTIFIC AMERICAN, INC

"I'D HAVE TO GO 30 YEARS WITHOUT MISSING A PUTT

"Let's face it, on most days, I'm lucky to go three holes without missing a putt. So when I say 30 years, I haven't gone off the deep end.

"I'm just making a point about the unbelievably high quality standards that Motorola Cellular sets for itself.

"The fact is, they're among the highest in the world.

"And since Motorola cellular phones are made so well, it should come as no surprise that they're the world's best sellers, delivering the kind of customer satisfaction nobody else can match.

"Motorola. It's an American company setting standards of quality the whole world can shoot for.

"Something that gets a big thumbs up from me."

TO HIT THE QUALITY STANDARDS MOTOROLA SETS FOR ITS CELLULAR PHONES."

MOTOROLA
Cellular Telephones

Technology that can take it.

© 1992 SCIENTIFIC AMERICAN, INC

Cells for Jerry's Kids

Experts argue the merits and safety of human trials

This past Labor Day on the Muscular Dystrophy Association (MDA) telethon, Jerry Lewis rendered his annual update of the encouraging progress being made against crippling illnesses. What viewers might not have guessed is that one area of therapeutic investigation—a form of cell transplantation called myoblast transfer—has sharply divided muscular dystrophy researchers. Although some regard myoblast transfer as a safe technique, others worry that experiments of dubious scientific value are exposing young patients to unwarranted hazards.

"People jumped too fast from animal models into humans," says Henry F. Epstein of the Baylor College of Medicine, a scientific adviser to the MDA. In a recent letter to *Science,* he and more than two dozen other muscle researchers and physicians called for a general moratorium on myoblast transfers in humans until certain basic questions have been resolved in animals.

"I truly believe their letter is propaganda by molecular geneticists intended to smear cell transplantation," huffs Peter K. Law of the Cell Therapy Research Foundation in Memphis. Law has reported the greatest successes with myoblast transfer—and his work has drawn the greatest fire. "They are opposed to it for fear that myoblast transfer will become the therapy instead of gene therapy," he says.

The disease at the center of the controversy is Duchenne's muscular dystrophy, a genetic disorder that strikes about one in every 3,500 boys and causes progressive wasting of the muscles. Those afflicted with the disease begin to weaken sometime after the age of five and gradually lose all strength in their limbs; they usually die by age 20 when their diaphragms or hearts fail. In the mid-1980s geneticists learned that Duchenne's dystrophy is caused by a defective gene for dystrophin, a protein essential to muscle function. Since then, investigators have sought to rescue the sick muscles by restoring the missing protein. Molecular biologists, for example, have been trying to develop a gene therapy that would insert working dystrophin genes into the muscles.

Myoblast transfer is an alternative approach in which whole cells, not just genes, are used. Experiments on rodents suggested that if cells called myoblasts are injected into dystrophic muscles,

© 1992 SCIENTIFIC AMERICAN, INC

they can sometimes save the unhealthy muscle fibers by fusing with them and making dystrophin. In 1989 George Karpati of the Montreal Neurological Institute, Law (who was then at the University of Tennessee at Memphis) and others initiated trials on human subjects. To date, more than 60 boys have received myoblast transfers.

Generally, the results of those experiments have been underwhelming: few of the subjects have demonstrated any improvement. This past March in *Cell Transplantation*, though, Law announced the best results yet. Using a patented technique, he and his colleagues injected myoblasts into the leg muscles of dystrophic boys and later measured their functional changes. Law's group claims that in the 13 boys tested so far, 81 percent of the treated muscles became stronger or did not lose strength.

Many muscular dystrophy researchers contend that Law's study is seriously flawed. Their primary complaint is that Law did not use a control group to test whether the gains are illusory. But Law insists that medical ethics prevented him from running the kinds of controls his critics wanted.

Law says he did not inject a placebo into patients because, in his experience, such injections accelerate muscle degeneration. He also did not inject myoblasts into just one leg of a boy and use the other leg as a control because the imbalance in strength might cause the boys to fall and hurt themselves. The only acceptable control, Law says, was to compare a muscle against itself after the myoblast injections. He likens the comparison to the before-and-after pictures used in commercials for baldness remedies.

"An experiment without controls is not an experiment," Epstein responds. He argues that pharmaceutical companies have established statistical methods for conducting drug trials that protect human subjects without invalidating the control procedures. Moreover, a control was critical in Law's experiments because some data suggest that cyclosporine, the immunosuppressive drug given to the boys to prevent them from rejecting the transplanted myoblasts, can cause a transient increase in the strength of muscles.

Law's is not the only work on myoblast transfer that has been criticized. This past April a team headed by Helen M. Blau of the Stanford University School of Medicine published a report in *Nature* confirming the ability of transplanted myoblasts to make dystrophin for sick muscles. Yet Eric P. Hoffman of the University of Pittsburgh

© 1992 SCIENTIFIC AMERICAN, INC

School of Medicine, one of the discoverers of dystrophin, argues that Blau's results may be less than they seem: a phenomenon called genetic reversion could cause sick muscle fibers to look as though they were making dystrophin. In any event, Hoffman says, no more than about 1 percent of the fibers seemed to have been helped by the treatment, an efficiency too low to have therapeutic meaning. "Is this really success?" he asks.

Blau stands by the conclusions of her paper. She emphasizes that, unlike Law's experiment, hers was a controlled, double-blind study; she also argues that her methods ruled out false-positive results from genetic reversion. Blau believes the petition for a moratorium is misguided but acknowledges that because the efficiency of the myoblast transfer was "surprisingly low," her group is deliberating about whether human experimentation "is the way to proceed at this time."

Defenders of myoblast transfer experiments point out that, according to the reports, no boy has ever been harmed by the procedure. Epstein is not reassured. "It's extremely inconsistent with all previous experience with cyclosporine that no side effects have been observed," he says, adding that those side effects include kidney failure and cancer. "We have no idea whether risks could be uncovered," he says. He and Hoffman both note that myoblast transfer is impractical for fixing what actually kills Duchenne's dystrophy patients—the degeneration of the heart and diaphragm.

The call for a moratorium, however earnest, carries no official weight. The Food and Drug Administration has not yet asserted any jurisdiction over cell transplantation. The National Institutes of Health are not currently funding any myoblast transfer experiments in humans. The MDA, which has funded experiments, remains open-minded on the subject. "The MDA is in the business of finding causes and cures," asserts Donald S. Wood, the MDA's director of science and technology. "The association would never cut off support to an area as long as it held promise."

In Wood's opinion, both human and animal experiments have value today. He sees the conflict over myoblast transfer as part of the "growing pains" all new technologies face. "It makes everyone work that much harder at getting the right answer," he says. But he also notes, "You must be humble in the face of this. You must go very cautiously. God forbid you should do anything to take even one day off the life of a child." —*John Rennie*

© 1992 SCIENTIFIC AMERICAN, INC

The world is at your fingertips with...

The Grundig World Band Receiver

Announcing a revolutionary breakthrough in world-band radios...from Grundig, the leader in German shortwave electronics. This miracle of miniaturization is not sold in U.S. stores or catalogs—only Willabee & Ward has it!

Tune in to events around the globe.

The RK-709 is the world's smallest 12-band receiver—AM, FM, FM Stereo (with headphones, included), and *all* 9 major shortwave bands. Small enough to fit into a pocket, briefcase, or purse, yet big on features. Telescoping antenna pulls signals from anywhere in the world. Listen to legendary BBC newscasts—the best in the world. Hear what Radio Moscow is saying about Yeltsin, or the American elections. Listen to the Voice of America for news from home when you travel abroad. Follow a crisis or breaking story wherever you are and no matter where it's happening: Baghdad, Barcelona or Beijing.

You don't have to pack a travel alarm, either.

You won't need it—this is better. The RK-709 wakes you with the radio or its ingenious "Humane Wake System"—the tone begins softly and gathers volume until you wake up! The digital alarm clock features a sleep timer that turns off the radio 10 to 60 minutes after you turn in!

Only $159!
Payable in monthly installments.

The Grundig RK-709 is only $159 (plus $7.50 shipping/handling), payable in convenient monthly credit card installments. Includes deluxe carrying pouch, three "AA" batteries, and stereo headphones. One year warranty, 30 day money back guarantee. Order today.

Don't be confused by other 4, 5, 7, or 9 band radios. Grundig's RK-709 is today's best buy on a real 12 band receiver.

CALL TOLL-FREE: 1-800-367-4534 Ext. 623-335

Willabee & Ward
47 Richards Avenue
Norwalk, Conn. 06857

W&W — Satisfaction guaranteed.

CALL TOLL-FREE: 1-800-367-4534
Ext. 623-335

Please send me ____ Grundig RK-709 World Band Receiver(s). For each receiver, charge 4 installments of $41.63* to my credit card:

☐ MasterCard ☐ VISA
☐ Discover ☐ Am. Ex.

Credit Card # ____ Exp. Date

Name ____

Address ____

City/State/Zip ____

Signature ____
(All orders subject to acceptance.)

☐ I prefer not to pay by credit card and will pay by check. Enclosed is my check for $159* plus $7.50 shipping/handling, a total of $166.50 for each receiver.

*Any applicable sales tax will be billed with shipment.

Phone orders normally shipped next business day.
Higher shipping/handling outside U.S.

© 1992 SCIENTIFIC AMERICAN, INC

Stellar Bells

Quivering stars bare their inner secrets

The discovery that the sun rings like a bell, made early in the 1960s at the California Institute of Technology, heralded the new field of helioseismology. Astronomers who watch oscillations of the solar surface can measure conditions deep within the sun, in the same way that seismologists monitor earthquake waves to study the interior of the earth. In the past few years, researchers have taken that remarkable achievement a step further: they are deducing the internal structures of distant stars from subtle vibrations on their surfaces, a technique called asteroseismology.

By far the greatest successes in this field have come from observations of white dwarf stars, the compact remnants of sunlike stars. Instabilities in the outer layers of some dwarfs set up waves that travel along a star's surface so that "the whole star shakes and shimmies," explains Steven D. Kawaler of Iowa State University. Those waves compress the outer layers of the star as they travel, causing certain regions to grow hotter and hence to radiate more intensely. Oscillations therefore show up as complicated but well-ordered changes in the brightnesses of white dwarfs. In the most extreme cases, a star's luminosity can vary by 30 percent.

One of the most crucial elements of stellar oscillation observation is that the record must be continuous. In 1988 R. Edward Nather of the University of Texas and a number of collaborators established the Whole Earth Telescope, a loose association of astronomers around the world dedicated to maintaining round-the-clock coverage of oscillating stars. The latest version of the Whole Earth Telescope, which began work on September 21, incorporates 13 observing sites, the largest number yet.

Some of the most impressive results from the enterprise concern an extremely hot white dwarf known only as PG 1159-035. A group led by D. E. Winget of the University of Texas reported that PG 1159-035 oscillates at 125 frequencies having periods from 385 to 1,000 seconds long. Hidden within those frequencies is a bounty of information about the physical conditions of the star.

The star's rotation, for example, causes oscillations that move west to east along the star's surface to exhibit a slightly different frequency than do oscillations moving from east to west. The magnitude of the frequency split indicates that the star completes a rotation every 1.38 days. That information provides clues regarding the evolution of red giants into white dwarfs. "This is the first piece of data on what the cores of red giants do," Kawaler says.

Asteroseismology is also illuminating other important aspects of stellar evolution theory by revealing the composition of white dwarfs. Places where a star's temperature, density or composition suddenly change act to reflect and trap internal waves. Seismologists exploited the same phenomenon to deduce that the earth is divided into a core, mantle and crust. Similarly, Winget and his colleagues succeeded in measuring the trapped oscillations to de-

Flotsam Footwear

Serendipity often comes to the aid of science. An amateur astronomer spots a nova, a fisherman captures a coelacanth in his net, a pair of oceanographers map the ocean currents by monitoring the advance of an accidental shoe spill....

A shoe spill? On May 27, 1990, a freighter was buffeted by a severe gale in the northeast Pacific Ocean, and five shipping containers of Nike footgear went over the side. Like a fleet of message-bearing bottles, the 80,000 sneakers began washing ashore in British Columbia, Washington and Oregon in early 1991.

When Curtis C. Ebbesmeyer of Evans-Hamilton, a marine instruments company in Seattle, and W. James Ingraham, Jr., of the National Marine Fisheries Service heard news reports of the shoe spill, they immediately realized they had stumbled across a potentially useful ocean drift experiment. "I tried to find the scientists who were tracking down the shoes, but nobody was," Ebbesmeyer recalls. "It surprised me."

Ebbesmeyer contacted his friend Ingraham, who monitors surface currents to determine their effects on salmon migration. With the eager assistance of a network of beachcombers, the researchers recovered about 1,300 of the shoes. Because the location of the spill was well known, the peripatetic shoes could provide a calibration point for computer models of ocean surface currents.

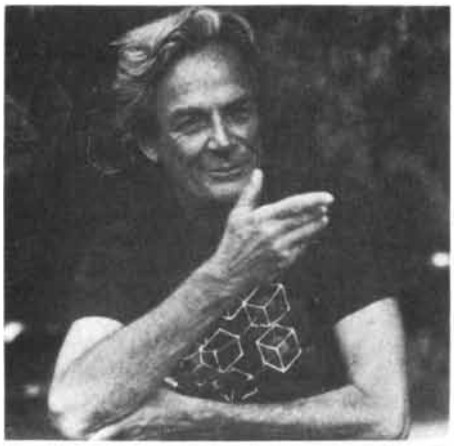
Ingraham then ran a computer hindcast to retrace the path of the shoes. It was "a perfect little get-together," as he describes it. His model showed that the 1990 path of drift was much farther south than usual. In certain other years, such as 1982, ocean currents associated with warm water in the tropical Pacific would have caused most of the shoes to drift toward Alaska.

The scientific value of the spill has by no means dried up. Some of the shoes recently reached Hawaii, and others "should be reaching Japan shortly," Ebbesmeyer notes. Any additional shoes that wash ashore will help the researchers as they expand their study of ocean surface drift to the western Pacific.

The great shoe spill of 1990 has also had some practical effects. Artist Steve McLeod of Oregon has earned $568 by collecting and selling the seafaring footwear. And both Ebbesmeyer and Ingraham are sporting their own recovered Nikes. Ebbesmeyer recommends giving the shoes a hot-water wash before wearing them; a long period of drifting may be good for science, but it is bad for comfort. "The shoes are real stiff after two years in the ocean," he reports. —*Corey S. Powell*

***FOOTLOOSE DRIFTER** recently washed ashore in California. Photo: Jackie Cunningham.*

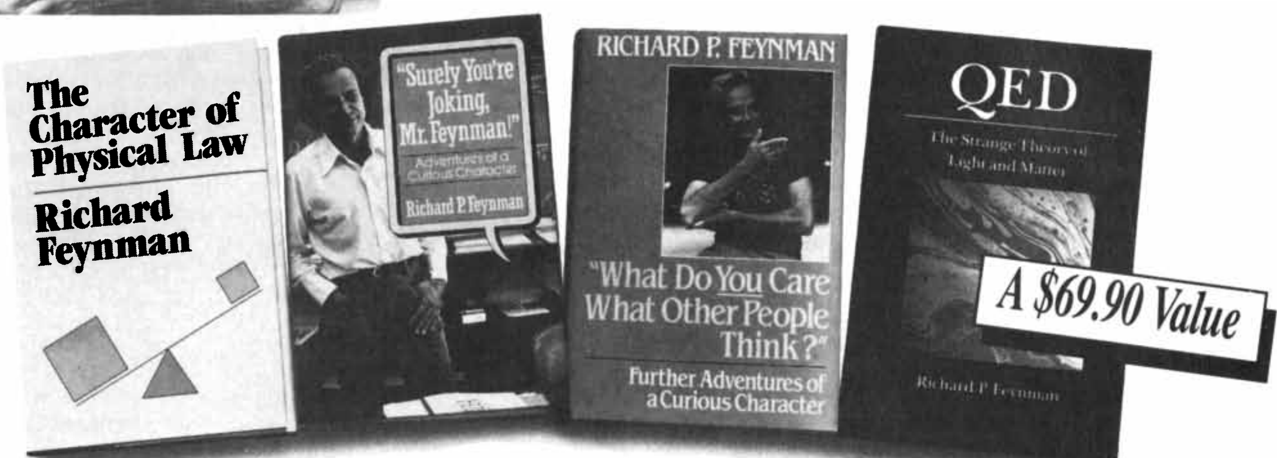
© 1992 SCIENTIFIC AMERICAN, INC



Take The Essential Feynman Collection for only $2.95

As your introduction to The Library of Science

You simply agree to buy 3 more books—at significant savings—within the next 12 months



You know his name. And you're probably familiar with his work. But who *was* Richard Feynman? What made him one of the most extraordinary physicists of his day?

Find out in *The Essential Feynman*. Here, in a superb four-volume collection, you'll find the complete man revealed. Best of all, for a limited time only, you can get this unique chronicle of Feynman's life, theories, and accomplishments for the special, low price of just $2.95.

QED
The Strange Theory of Light and Matter

Feynman's work in quantum electrodynamics earned him a Nobel Prize. Now, you can explore this complex topic in a presentation remarkable for its clarity, completeness, and humor. Based on a famous lecture series at UCLA, this book presents Feynman at his best: describing the interaction of light and physics ... unraveling the phenomenon of iridescence ... relating how quantum electrodynamics helps us understand quarks, gluons, and other major aspects of contemporary physics. *Publisher's Price: $18.50*

The Character of Physical Law

Undoubtedly a brilliant scientist, Richard Feynman was also a dynamic lecturer. This collection of transcripts from a 1964 lecture series at Cornell University is *classic* Feynman. His enthusiasm is unbridled as he presents an overview of physical laws ... expounds on the elegance of scientific law ... demonstrates the interaction of mathematics and physics ... and much more. Don't miss one of Feynman's most insightful, entertaining performances, in a hardcover edition brought back into print exclusively for The Library of Science.
Publisher's Price: $17.50

"Surely You're Joking, Mr. Feynman!"
Adventures of a Curious Character

Cracking safes at Los Alamos. Accompanying ballet on a bongo drum. Explaining physics to Einstein. It's *all* essential Feynman, a man as renowned for his antics as his accomplishments. And it's here for you to enjoy in memoirs recorded by crony Ralph Leighton.
Publisher's Price: $15.95

"What Do You Care What Other People Think?"
Further Adventures of a Curious Character

Once, when discussing the Challenger space shuttle disaster, Richard Feynman stated, "Reality must have precedence over public relations ... Nature cannot be fooled." Succinct and telling, these words describe the man whose interests, philosophies and, in particular, experiences as a member of the Presidential Commission investigating the Challenger explosion, are brought to life before your eyes. Written with long-time friend Ralph Leighton, this book is the legendary physicist's last literary legacy.
Publisher's Price: $17.95

THE LIBRARY OF SCIENCE is the nation's oldest and most respected science book club. Dedicated to serving both professionals and serious amateurs, it surveys the entire field of published sources on scientific subjects to make the very finest works available to club members at substantial savings, with prices starting as low as $9.95.

MEMBERSHIP BENEFITS- In addition to getting the 3-volume Essential Feynman Collection for only $2.95 when you join, you'll also receive discounts on books you choose to purchase. Discounts generally range from 15% to 30% off the publishers' prices, and occasionally even more. • Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of more than 50% off the publishers' prices. • Moreover, our books are always identical to the publisher's editions. You'll never receive an "economy edition" with inferior paper and bindings from us. • At 3-4 week intervals (15 times per year), you will receive the Library of Science Book Club News, describing the coming Main Selection and Alternate Selections, together with a dated reply card. • In addition, up to three times a year, you may receive offers of Special Selections which will be made available to a group of select members. • If you want the Main Selection, do nothing, and it will be sent to you automatically. • If you prefer another selection, or no book at all, simply indicate your choice on the card and return it by the date specified. • You will have at least 10 days to decide. If, because of late mail delivery of the News, you should receive a book you do not want, we guarantee return postage.

If reply card is missing, please write to Library of Science, Dept. 2-DP9-00954, 3000 Cindel Drive, Delran, NJ 08370-0001, for membership information and an application.

Scientific American 11/92

© 1992 SCIENTIFIC AMERICAN, INC

termine the interior structure of PG 1159-035. It seems to have a three-layer composition: a blanket of hydrogen atop helium and carbon shells.

More extensive observations will enable the researchers to deduce not only the density of the various layers but their thicknesses as well. When applied to a whole range of white dwarfs, that technique will make it possible to refine significantly the theory of the internal transformations that occur as stars evolve.

Another noteworthy application of asteroseismology is indicating the rate at which white dwarfs cool. Such information "will tell you how old the oldest white dwarfs are—that's our ultimate goal," Nather says. Knowing the ages of the oldest dwarfs will reveal the minimum age of the Milky Way and also may make it possible to learn whether the galaxy formed all at once or if various parts coalesced at substantially different times.

According to theoretical models, the coolest known white dwarfs seem to be about 10 billion years old, another piece of evidence that the universe cannot be only eight billion years old, as some cosmological observations imply. Kawaler reports that the observed rate of change in the oscillation periods caused by cooling is "on the right time scale" expected for hot white dwarfs. He is still studying the signs of change on cooler stars—a rather difficult task, since he is looking for an effect amounting to "a few seconds over 15 years of data," he notes. So far asteroseismology shows that the theory is "right to a factor of two."

White dwarfs are particularly open to the techniques of asteroseismology, but astronomers are looking for oscillations among all kinds of stellar populations. Ideally, one would like to analyze the vibrations of stable, middle-aged stars and compare their properties with those of the sun. Oscillations alter the brightness of such stars by only a few parts per million, however. A group led by Timothy M. Brown of the High Altitude Observatory tentatively identified oscillations on the star Procyon, but he readily confesses that "there are a frightening number of ways you can get fooled." His group will make another set of observations this winter to try to improve their level of certainty.

Meanwhile nobody is forgetting about the most proximate star. The Global Oscillation Network Group, an international grouping of six telescopes, will provide several years of continuous observations of solar oscillations starting in 1993. Helioseismology is already answering such seemingly imponderable questions as, What is the temperature at the core of the sun, and what drives the 22-year activity cycle? Nather reflects philosophically on this kind of work: "It's amazing that nature would allow us to do it, that it's possible at all." *—Corey S. Powell*

Who Were the Indo-Europeans?

Did successive waves of Indo-European horsemen begin to gallop from the steppes of southern Russia some 6,500 years ago, spreading their language as they subjugated farmers between Greece and the Ganges? Or did the vast Indo-European family of languages expand from the Middle East beginning 2,500 years earlier, when farmers moved outward in search of land, swamping any foraging cultures in their path?

Rather than resolving this prehistoric puzzle, a recent study of European genetic patterns by Robert R. Sokal and his colleagues at the State University of New York at Stony Brook has failed to verify either theory. Neither can account for the observed correlation between the languages and genes of Europe.

Sokal's team set out to test an earlier model that Albert J. Ammerman of the University of Parma and L. L. Cavalli-Sforza of Stanford University had proposed to explain the spread of agriculture from the Fertile Crescent. The researchers argued that a genetic trend between the southeastern and northwestern extremes of Europe was the vestige of a population boom occasioned by the invention of agriculture in the Fertile Crescent. In this view, farming had been propagated less by cultural borrowing than by demographic replacement. In 1987 Colin Renfrew of the University of Cambridge applied the model to his analysis of the archaeological record to explain the spread of Indo-European languages.

Last year the Stony Brook researchers mapped European genetic patterns against the archaeological record of early farming cultures and confirmed the demographic leg of the theory. At the end of the study, published in *Nature,* the workers cautioned that further research was necessary to test the linguistic leg, which posits 10 postagricultural transitional areas where the Indo-European subfamilies differentiated into their present form.

Now Sokal's latest study, published in the *Proceedings of the National Academy of Sciences,* leaves that leg hanging in thin air, right next to the warrior theory of Marija Gimbutas, an archaeologist at the University of California at Los Angeles. This time Sokal and his team analyzed the genetic patterns of many different European populations so they could fit them into a family tree. Then they compared that tree with one of 43 European languages assembled by Merritt Ruhlen, a linguistic taxonomist. The workers found a 0.14 correlation between genes and languages.

The researchers then estimated how much of that correlation could be explained by mere geographic distance, a factor that differentiates genes and languages in tandem. After holding geography constant, the workers found an average residual correlation between languages and genes of 0.06. This residuum remains unexplained.

Sokal demonstrates the crux of the experiment by superimposing first Renfrew's theory, then Gimbutas's, on a map of Europe. Each theory appears as a set of arrows, placed with the help of Renfrew and Gimbutas. "If Renfrew's theory were true," Sokal says, "it should account for the rest of the correlation, which should then drop to zero. But there's no change. It's still 0.06. The same is true when you add Gimbutas—0.06 correlation."

Such small correlations may seem mere statistical static, but Sokal insists they are significant. "Not every genetic locus will differentiate during the origins of various populations," he declares. "In a comparison of modern, racially diverse populations—Italians, Nigerians and Japanese—Italians differed from the other two populations by as much as 0.2 in only 20.4 percent of the cases."

How, then, does Sokal account for the correlation? He declines to offer a model but says he has an inkling of what might have happened. If geography cannot explain the concurrence of languages and genes, then both must have begun to evolve in parallel in some other place, perhaps outside of Europe. In that case, they must have come with immigrants as yet unknown. *—Philip E. Ross*

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

A Macintosh to start a

Imagine you could design your own personal computer. What would it be? It would be fast, of course. Faster than your basic 486 machine. Fast

The Apple Macintosh Quadra.

enough to handle the rendering, animating, design and analysis jobs that often require a workstation.

It would have immense storage and memory capacity. And it would have features like high-performance networking and accelerated video support built in.

Macintosh Quadra runs all the most powerful engineering software.

You would make it flexible enough to run any kind of application. Affordable enough to put on every engineer's desk. And as long as you're fantasizing here, you'd make it as easy to set up, easy to learn and easy to use as an Apple® Macintosh® personal computer.

But it's not a fantasy. It's a Macintosh Quadra.™

By any measure, the Macintosh Quadra 700 and 950 are two of the most powerful personal computers ever built. Both are based on the Motorola 68040 (rated at 20 and 25 MIPS, running at 25 and 33 MHz), which integrates the processor, math coprocessor and RAM cache all onto one chip.

They're up to twice as fast as any of their forebears. Fast enough to beat the chips off comparably priced 486 computers from IBM, Compaq and Dell.* And fast enough to make programs like AutoCAD, MicroStation Mac,® MacBRAVO! and VersaCAD perform at a level once seen only in dauntingly complex workstations.


Built-in 24-bit video lets you create photo-realistic images without the expense of an extra card.

High-performance subsystems provide built-in support across the board: Ethernet networking, accelerated 24-bit video** support and faster SCSI and NuBus™ slots.

You can add a 400MB hard disk to both

*Comparisons are based on a 1991 independent research study conducted by Ingram Laboratories that tested a variety of personal computers running applications available for both the Macintosh and Microsoft Windows 3.0 environments. **24-bit video support for up to a 16" monitor specifications and descriptions were supplied by the respective vendor or supplier. Orbiter images created by Jerry Flynn. ©1992 Apple Computer, Inc. Apple, the Apple logo, A/UX, Mac, Macintosh and "The power to be your best" are registered trademarks and Macintosh Quadra and SuperDrive are trademark of Wolfram Research, Inc. MicroStation is a registered trademark of Bentley Systems Inc. MS-DOS is a registered trademark and Windows is a trademark of Microsoft Corp. Motorola is a registered trademark of Motorola Corp. NuBus is a trademark of Texas Instruments.

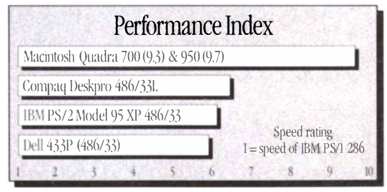
© 1992 SCIENTIFIC AMERICAN, INC

with the power revolution.

communications architecture lets you easily take advantage of features like distributed processing, allowing programs like RenderPro and BackBurner to utilize excess CPU cycles on other Mac computers or workstations for faster renderings.

There are two Macintosh Quadra models. The 950 sits next to your desk, the 700 fits on top of it.

And Macintosh Quadra fits in with the PCs you already own. The built-in Apple SuperDrive,™ used with Apple's new Macintosh PC Exchange software, allows you to share files easily with MS-DOS PCs via floppy disk or over a network. And with Apple's fully compliant version of UNIX®—A/UX®—you can even run UNIX, X Window, MS-DOS and Macintosh programs all at the same time.



Macintosh Quadra significantly outperforms 486 PCs from IBM, Compaq and Dell running Windows 3.0.

It all adds up to the kind of power that moved *PC Week* to give the Macintosh Quadra its highest satisfaction rating in the categories of overall performance, price relative to performance, expansion capability and ease of installation and configuration.†

the 700 and the 950 to accommodate the largest CAD files. And the 950 even lets you add a CD-ROM drive or a removable cartridge drive, and a disk array or more than a gigabyte of internal hard disk storage.

You can increase the memory of the 950 to up to 64MB of RAM for handling compute-intensive applications like three-dimensional modeling and stress analysis.

Over your network, the unique Apple interapplication

Your authorized Apple reseller would be glad to show you all this and more. For the name and location of the one nearest you, call 800-538-9696, extension 200.

And soon you'll discover the latest power of Macintosh personal computers. The power to start a revolution. The power to be your best.®

available for 21" monitors with optional third-party video card. †PC Week, December 23/30, 1991. ©1991 Ziff Communications Company. Mention of third-party products is for informational purposes only and constitutes neither an endorsement nor a recommendation. All product ...marks of Apple Computer, Inc. AutoCAD is a registered trademark of Autodesk, Inc. BackBurner is a trademark of Specular International. MacBRAVO! is a trademark of Schlumberger Technologies, Inc. MacTOPAS is a registered trademark of AT&T Graphics Software Labs. Mathematica is a ...rPro is a trademark of Strata, Inc. Vellum is a registered trademark of Ashlar Incorporated. VersaCAD is a registered trademark of Versacad Corp. X Window System is a trademark of MIT. UNIX is a registered trademark of AT&T. This ad was created using Macintosh computers.

© 1992 SCIENTIFIC AMERICAN, INC

Desert Dynamics

Competition is the rule in complex ecosystems

To anyone who has watched birds jostling one another at a garden feeder, the idea that species compete for valuable resources might seem obvious. But strange to say, ecologists have often disagreed about how important competition actually is in natural ecosystems. Some researchers argue that climatic factors such as temperature and the amount of rainfall are likely to be far more critical. Clashes over food, according to this way of thinking, are significant only during hard times, when there is a shortage of alternative foraging places.

One reason the debate about competition has gone on for so long is that counting animals in natural or near-natural habitats is notoriously difficult. And because field experiments are usually done with grants that last for only a year or two, most studies have been short. A study run by James H. Brown, a professor at the University of New Mexico, is one of the few exceptions. Brown is conducting a long-term investigation of the interactions between rodents, birds and plants in the Chihuahuan desert of southeastern Arizona. In a recent report, Brown concludes there is a persistent and steady competition between species despite the importance of climatic effects on the numbers of individuals.

Brown and his collaborators started their experiments 15 years ago by fencing off 24 plots, each 50 meters along a side, in flat desert near Portal, Ariz. The fences, made of wire mesh, extend 60 centimeters above ground and 20 centimeters below, to discourage the kangaroo rats, deer mice and other rodents being studied from burrowing under them. A few plots were left undisturbed. In some the experimenters excluded larger rodents by making holes in the fences that were too small for them to get through. In other plots they removed different combinations of species. Then the workers carefully documented the numbers of various plants and animals in the plots and watched how they changed over time.

In the plots that Brown and his colleagues left alone, the eight desert rodents that they studied varied strikingly in their responses to changing environmental conditions. Some species displayed a five-year repeated pattern that Brown links to the El Niño Southern Oscillation, a climatic cycle that causes heavy winter rain in the southwestern U.S. in some years. But other species showed no effect—even though all the rodents feed on seeds, which plants produce in greater numbers in wet years than in dry years.

The studies indicate that it would have been impossible to predict how each species would respond. Brown notes that his experiments provide no support for the idea—beloved of biology textbooks—that ecological communities reach equilibria appropriate to their geographic region. Nature is not so dull. In his experimental plots, equilibrium was the only thing that was never present. Rather, he says, "community composition varied continuously over time."

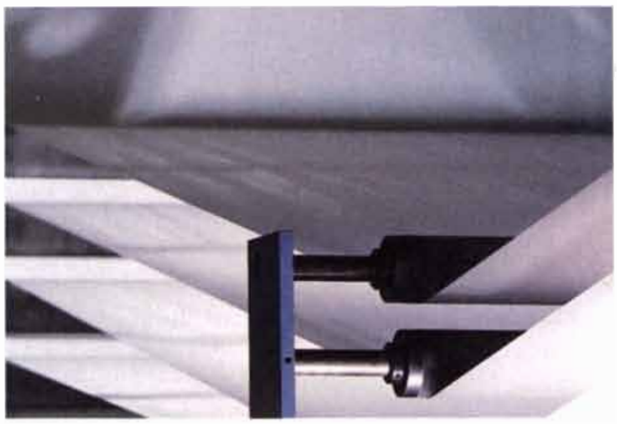
The experimental removals of particular species from plots provided further evidence of complex dynamics even in a relatively simple ecosystem. Removal of a species, Brown says, can lead to cascading effects that take years to play out. When he removed the three species of kangaroo rats from several plots, for example, he found that the habitat changed dramatically over the course of a few years. Several species of grasses colonized the areas between shrubs and became far more abundant while other, short grasses became rare.

Birds, like rodents, forage on seeds, so it would not be surprising if removing the rodents from a plot made it more attractive to birds. But in fact

***VEGETATION CHANGES** markedly when kangaroo rats are removed from an experimental plot* **(to left of fence)**. *After five years, the annual* **Lesquerella gordonii (yellow flowers)** *is more common* **(top)**. *After a further eight years, tall perennial and annual grasses dominate* **(bottom)**. *Photos: James H. Brown.*

© 1992 SCIENTIFIC AMERICAN, INC

Electricity Can Help Increase America's Productivity Without A Lot Of Energy.


Electric ultraviolet technologies expedite the drying of ink on paper.

Productivity is key to prosperity. That's why America's electric utilities are using their resources to help find new ways to maximize speed and efficiency, while enhancing quality.

For example, food producers gain in speed and quality from electric microwave drying processes.

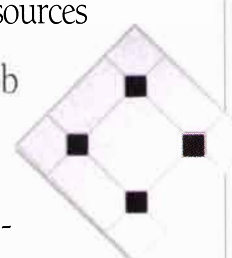

Producers of such foods as pasta gain in speed and quality from electric microwave drying processes.

Printers apply electric ultraviolet technologies to expedite the drying of ink on paper. Electric induction heating processes speed up the manufacture of metal and glass products.

Everywhere, electric technologies are boosting our nation's productivity.

In most cases, speed and quality are improved at lower costs and with less emissions. And, in all cases, with a considerable savings in energy.

American industry is committed to stepping up productivity. The electric utility industry shares this commitment. We will continue to devote our resources to getting the job done cleanly, quickly and efficiently. For more information, contact the Edison Electric Institute at 1-800-438-8334 or, in Washington, at (202) 508-5002.

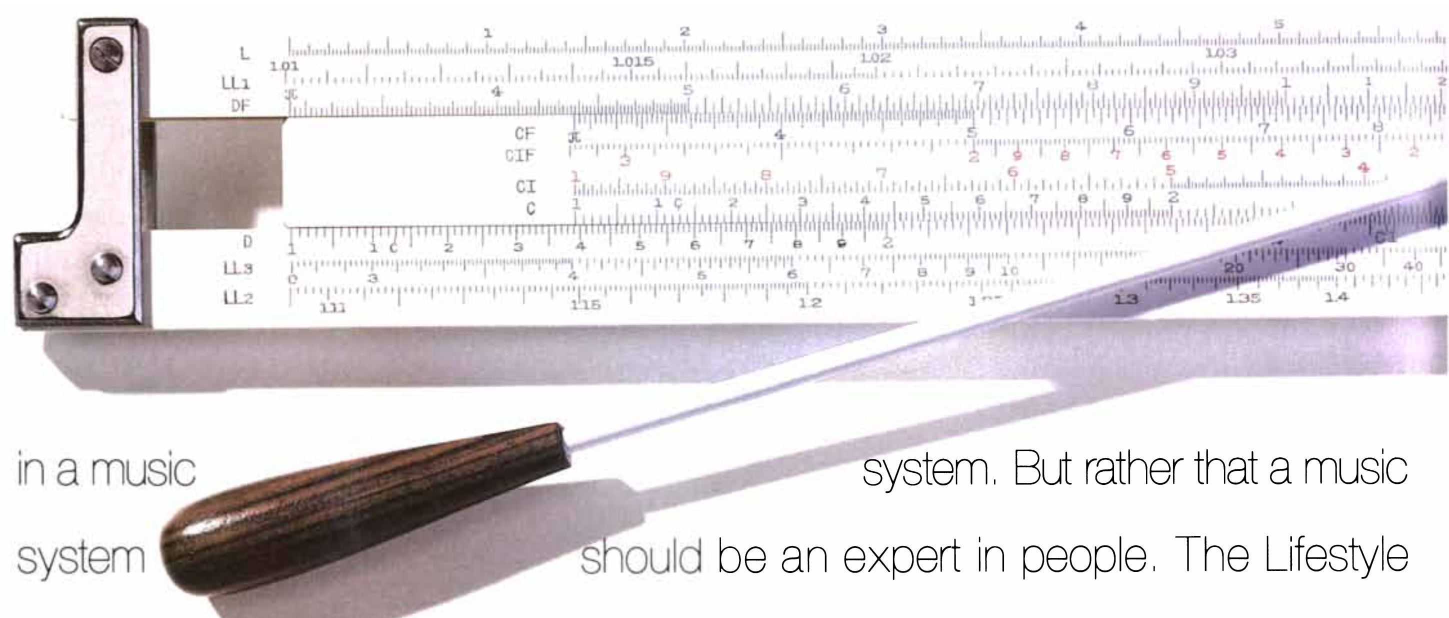

The building industry uses electric technology to harden the coats of no-wax floors.

ELECTRICITY
Taking you into the future.

© 1992 SCIENTIFIC AMERICAN, INC

M.I.T. meets Mozart.

This is the place where technology and art intersect. The kind of thinking you'd expect from engineers who went to M.I.T. But in their former lives probably played under the baton of Wolfgang Amadeus Mozart. We call it the Bose® Lifestyle® music system. You'll find it is nothing less than a whole new way of listening to and enjoying music. It is based on the idea that people shouldn't have to be experts

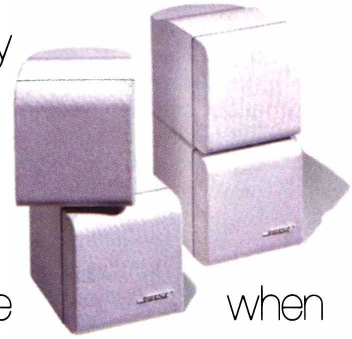


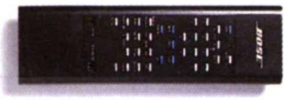
in a music system. But rather that a music system should be an expert in people. The Lifestyle music system is not complex. It is simple. It does not have hundreds of buttons. It has just seven. It does not take hours to set up. It takes minutes. It is not big. It is compact. It is "the most thoughtful simplification of audio design" says

Hans Fantel. In the New York Times, he called our CD player/AM/FM tuner an

© 1992 SCIENTIFIC AMERICAN, INC

"exquisite Bauhaus-style jewel box." A metaphor no doubt inspired by the singular richness of the Bose sound. You see, for all its elegant simplicity it gives up nothing in terms of performance. At Bose, all the research, all the engineering, all the obsessive attention to details have but only one purpose: better sound. Because when all is said and done, a stereo system should not be measured by a slide rule, but by what happens to the short hairs on the back of your neck when you hear your favorite aria. Or blistering guitar solo. And in this regard, the Bose Lifestyle music system is unparalleled. At any price. Of course, there are people who would have you believe that one high-quality stereo system sounds pretty much the same as another. That may be true. Unless one of them happens to be a Bose. Call us today at the number below and you can receive it tomorrow via Federal Express® delivery service. Try it out in your living room for two weeks with absolutely no obligation. Once you experience what happens when art and technology intersect, you won't be satisfied with anything less than a Bose.





Call 1 800 645 BOSE, ext. 230 to order or to get more info on the Bose System.

BOSE®
Better sound through research™

© 1992 SCIENTIFIC AMERICAN, INC

the opposite happens, Brown's studies show. Birds forage less in plots that have no rodents than they do in undisturbed plots. Moreover, birds show a similar aversion to plots that have had all the ants removed—even though ants, like birds and rodents, forage on seeds. Brown and his collaborators think rodents and ants make plots more attractive to birds by making trails and creating areas of bare soil. Birds apparently avoided plots with denser vegetation because they would have to expend more effort to find the same amount of food.

Yet despite the complexity, a sophisticated statistical analysis proved that competitive interactions are still an important force. When Brown, together with Mark L. Taper and Edward J. Heske, examined the collective effects of kangaroo rats on the other seed-eating rodents in his plots over a long period, a clear signal emerged. The smaller rodents consistently forage by preference where the larger kangaroo rats do not. And the effect was remarkably constant over time.

Brown displays proper scientific caution about generalizing from his results. The Arizona desert might not be typical of other ecosystems, for example. Nevertheless, he notes, "our experiment is virtually unique in being of sufficient duration to assess long-term temporal variation in competition." And in a paper submitted to the journal *Ecology,* entitled "Constant Competition in a Variable Environment," Brown's conclusion about the role of competition is unambiguous: "Interspecific competition plays a major, sustained role in the structure and dynamics of this desert rodent community." —*Tim Beardsley*

Kicking Chaos out of Lasers

By reputation, lasers emit "pure" light of homogeneous wavelengths and consistent intensity. In fact, the intensity of light produced by some lasers often develops chaotic fluctuations. At the Georgia Institute of Technology, Rajarshi Roy and his graduate students Zelda Gills and Christina Iwata are finding ways to tame lasers that are chaotically inclined. In the process, they have found they can increase the stable power output of certain lasers by 15-fold.

The researchers work with a popular laser—a solid-state neodymium-YAG laser, pumped, or powered, by another diode laser. For several years, workers have used such lasers to produce short-wavelength green light by doubling the frequency (and halving the wavelength) of an infrared laser with the help of a nonlinear "doubling" crystal.

Such frequency doubling is nonetheless plagued by the so-called green problem: as infrared is converted to green, the intensity of the light spontaneously degenerates into chaotic oscillations. Trying to boost the output of light by pumping in more energy also triggers chaos.

Two years ago Roy and his team found they could dampen such chaos by carefully orienting the doubling crystal, thereby skirting the odd polarization effects that caused the green problem. But what if parameters changed?

INTENSITY OF LASER, shown on oscilloscope screen as chaotic bursts over time, is monitored by Georgia Institute of Technology graduate student Zelda Gills. Photo: Margaret Barrett, Georgia Institute of Technology.

As the researchers began looking for dynamic controls to stabilize chaos, they turned to a novel analog technique called occasional proportional feedback control. Starting with a laser that emits light that fluctuates chaotically in intensity, they sampled the signal at regular intervals. The differences between those measured values and a collection of reference values were then translated into tiny kicks that nudged the signal into periodic behavior.

The control was successful—to a degree. Once the Georgia Tech researchers tried boosting the power input into their laser, they provoked another onset of chaos. Help was at hand, however, in the form of recent theoretical work by Ira B. Schwartz and Ioana A. Triandaf of the U.S. Naval Research Laboratory aimed at controlling unstable orbits.

By using error-correcting codes, Schwartz and Triandaf had enabled algorithms that control local spurts of chaos to handle a wide range of changing parameter values. As a result, by tuning only one, easily accessible parameter (say, power input), the mathematicians could track the behavior of unstable orbits. Such an approach could consequently be used to compensate for random drift.

By integrating the error-correcting codes into their control program, Roy and his students managed to squeeze out 15 times more stable light than the laser had previously produced for minutes at a time. The controls, moreover, required little additional energy—only 2 or 3 percent of the pumping power.

"The laser experiment shows off the real power of applying mathematics to nonlinear systems," observes Schwartz, who has already filed a patent application. He is looking forward to trying out the approach in other arenas, including orbiting satellite platforms, fluid and combustion control systems and cardiac pacemakers. Roy, on the other hand, will continue to explore chaos in other laser and fiber-optic systems. Uncovering new sources of chaotic behavior to tackle is not yet a problem. —*Elizabeth Corcoran*

© 1992 SCIENTIFIC AMERICAN, INC

IT'S NOT SO FAR EAST ANYMORE.

NORTHWEST JUST BROUGHT ASIA HOURS CLOSER. At Northwest, we know people prefer to get to their business destination sooner rather than later. So we figured we'd knock several hours off our flights to many Far East destinations, giving you more time once you've arrived.

Northwest to Asia. *Fastest Pacific service to key business destinations.**	
New York/JFK to Hong Kong	*1 hr 10 min faster*
Boston to Manila	*1 hr 25 min faster*
Chicago to Bangkok	*1 hr 30 min faster*
Seattle/Tacoma to Singapore	*2 hr 55 min faster*
Northwest. The leading airline to Asia. Now with the fastest service of any airline to many Far East cities.	

And now we're faster to many Asian destinations than any other major airline. In fact, we've reduced our transit times through Tokyo by up to three hours. That's up to three extra hours to prep for a meeting, take a nap or do anything else you might want to do.

As usual, Northwest offers the most nonstop flights from the U.S. to Tokyo, Seoul and Osaka. And while you're in the air with us, you'll be treated to a host of new amenities, like Executive Class mini-sleeper seats, enhanced meal service and improved audio and video programming. All of which are good reasons why Northwest is the leading U.S. airline to Asia. So the next time you're flying to one of our 15 Asian destinations, fly the airline that shortened the trip.

NORTHWEST AIRLINES
Some People Just Know How To Fly.

CALL YOUR TRAVEL AGENT OR NORTHWEST
1-800-447-4747

*Indicates the difference between scheduled flight times for Northwest and the carrier with the next fastest scheduled flight, for travel between October 1, 1992 and October 8, 1992. Times are based on published schedules as of September 18, 1992 for all carriers scheduled from point of origin to destination, in its entirety. Actual flight times may vary and schedules are subject to change.

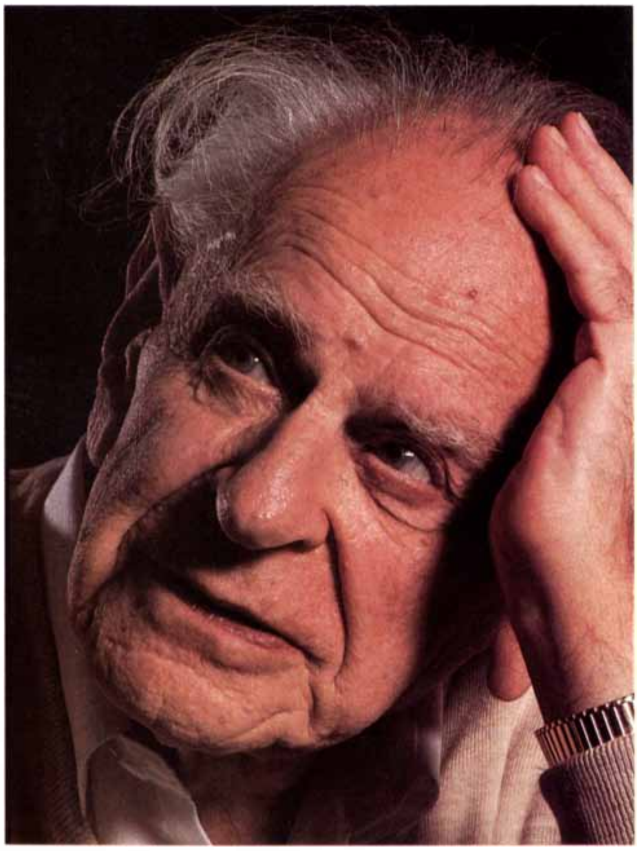
© 1992 SCIENTIFIC AMERICAN, INC

The Intellectual Warrior

I am just about to meet the philosopher Karl R. Popper, and I'm trying to lower my expectations. Popper is far and away the most influential philosopher of modern science—among scientists if not other philosophers. He is best known for his assertion that scientific theories can never be proved through experimental tests but only disproved, or "falsified." In countless articles and more than a dozen books—the latest, a collection of essays, published just this year—he has also held forth on quantum mechanics, determinism, the theory of evolution, political totalitarianism and practically every other issue of note.

But as Popper's assistant ushers me into his house south of London, she warns me that "Sir Karl" (he was knighted in 1965) is exhausted and will probably only have the energy to talk for an hour or so. He just turned 90 in July, a month ago, and he has already endured endless interviews and congratulations. For the past several days, he has been toiling over two lectures he intends to deliver when he receives the prestigious Kyoto Prize, also called Japan's Nobel, in November. On top of that he has been ill and is still taking medication.

Then Popper makes his entrance. Stooped, equipped with a hearing aid and surprisingly short (I'd assumed the author of such autocratic prose would be tall), he is nonetheless as kinetic as a bantamweight boxer. He brandishes an article I wrote for *Scientific American* about how quantum mechanics is inspiring some physicists to abandon their view of subatomic particles as wholly objective entities. "I don't believe a word of it," he declares in a Viennese-accented voice. "Subjectivism" has no place in physics, quantum or otherwise, he informs me. "Physics," he exclaims, grabbing a book from a table and slamming it down, "is that!"



AUSTRIAN-BORN PHILOSOPHER Karl R. Popper has inveighed against dogmatism in both science and politics throughout his career. Photo: David Levenson/Black Star.

Once seated, he keeps darting away to forage for books or articles that can buttress a point. Striving to dredge a name or date from his memory, he kneads his temples and grimaces as if in agony. At one point, when the word "mutation" briefly eludes him, he slaps his forehead with alarming force, shouting, "Terms, terms, terms!"

During one of his brief pauses for breath, I mention his assertion that a theory must be falsifiable to be considered scientific. Is this falsifiability theory, I ask him, falsifiable? Popper places his hand over mine and transfixes me with a radiant smile. "I don't want to hurt you," he says, his voice softening, "but it's a silly question."

Still smiling, he gently explains that "the function of falsifiability is to say whether a theory is scientific or not. My theory of falsifiability is certainly not scientific. It belongs to metascience." Popper used to banish students from his seminar for asking such an "idiotic" question, but he doesn't blame me for doing so; some other philosopher, he suggests, probably put me up to it. "Yes," I lie.

I should have known better than to try to trip up Karl Popper. For more than 70 years, he has been debating this century's greatest ideas with this century's greatest minds. And criticism, after all, is Popper's credo. He sees criticism, and even conflict, as essential for progress of all kinds. Just as scientists approach the truth through what he calls "conjecture and refutation," so do species evolve through competition and societies through political debate. A "human society without conflict," he once wrote, "would be a society not of friends but of ants."

Popper was raised in Vienna in an intellectual household; his father was a professor of law and his mother an accomplished musician. He traces the genesis of his philosophy, which he calls critical rationalism, to his 17th year. After a brief dalliance with communism, he became disgusted by the dogmatism of Marxists, by their utter certainty that Marx was "right." At roughly the same time, he learned that observations of a recent solar eclipse had borne out a prediction of a bizarre theory of gravity proposed by a young physicist named Albert Einstein.

The contrast compelled Popper to wonder: What exactly distinguished pseudoscientific theories, such as Marxism or astrology or even psychoanalysis, from scientific ones, such as Einstein's theory of relativity? The answer, he decided, was that the latter offered predictions specific enough to be experimentally tested—and hence falsified.

At the time, the philosophy of science was dominated by logical positivism, which asserted that scientists can logically infer certain limited truths about

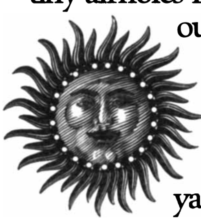
© 1992 SCIENTIFIC AMERICAN, INC

A shirt for those who prefer their logos on the inside.

The only place you will find an emblem on a Lands' End Mesh Knit Shirt is sewn quietly within the neckline.

What you will notice is how soft and unbelievably comfortable our pure combed cotton feels on you. This particular long-staple cotton is particularly smooth next to your skin.

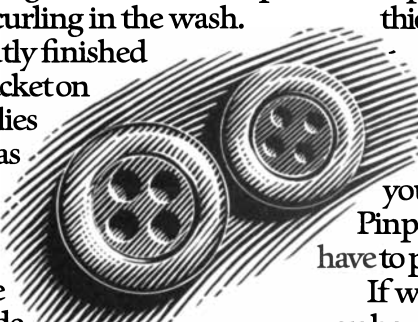
Of course, a sturdy mesh knit stitch might have something to do with it, too. Thousands of tiny airholes in the fabric help our shirt breathe, so you stay cool and dry. Even when it's hot outside.



A very unique yarn compacting process assures shrinkage is kept to a minimum. (About 7% in case you're one of those people who are into percentages.)

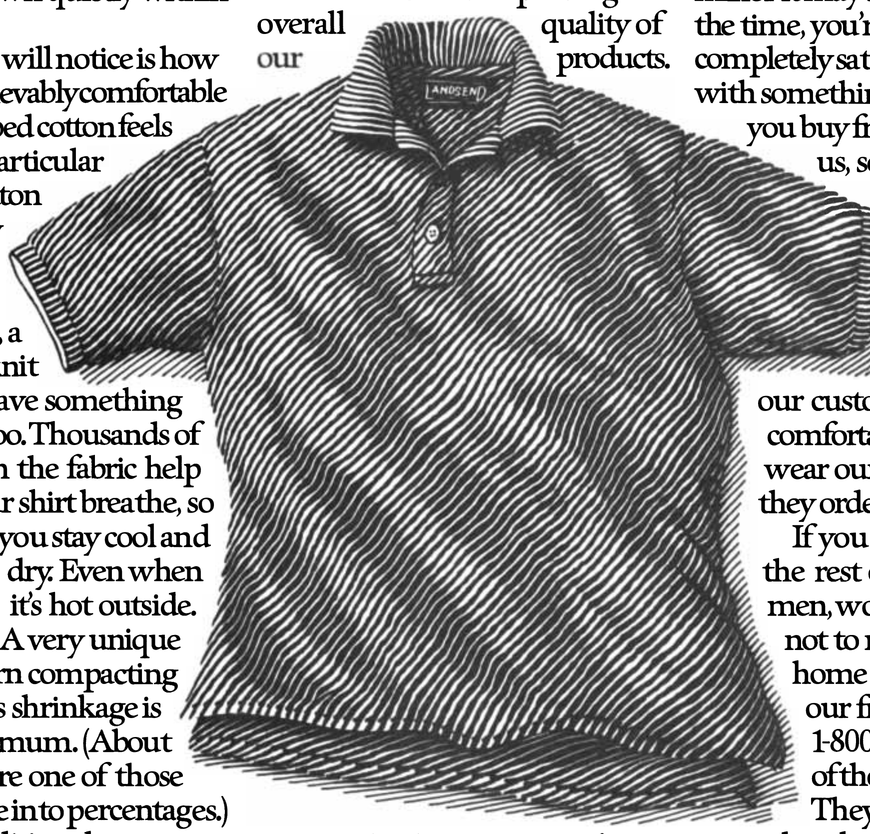
For an additional measure of comfort, we tape the shoulder and neck seams. While our knit collar is designed to hold its shape without curling in the wash.

A neatly finished classic placket on the front lies flat and has two cross stitched buttons.



There are also side vents and long tennis-style tails so you can wear it untucked, should that be your preference.

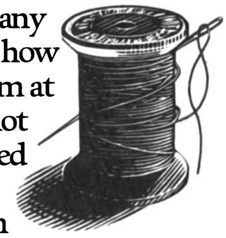
As with everything with our label on it, the value of our mesh shirt lies in the attention we pay to details. We're constantly adding new features and improving the overall quality of our products.



The Lands' End Mesh Knit Shirt, under $20.

Whether it's a timeless cotton Drifter Sweater, some weathered canvas pants or a complete set of thick Supima Towels for the bathroom, we make sure that every item lives up to our rigid standards and, more important, your expectations. (Our basic Pinpoint Oxford Shirts, alone, have to pass 71 tough inspections.)

If we have not yet convinced you how strongly we believe in our fine merchandise, it is all Guaranteed. Period.® Which means exactly that.



Simply put, if for any reason, no matter how minor it may seem at the time, you're not completely satisfied with something you buy from us, send it back. Pronto.

What you will get in return is all your money or another item.

It's our way of making sure that our customers are not only comfortable when they wear our clothes, but when they order them, too.

If you would like to view the rest of our clothing for men, women and children, not to mention our classic home furnishings, phone our friendly operators at 1-800-356-4444, any day of the week.

They'll be more than happy to send you a free Lands' End catalog with your name written all over it.

To get our free catalog call 1-800-356-4444.

Or fill out the coupon and mail it to:
1 Lands' End Lane, Dept. JH, Dodgeville, WI 53595

Name ______________________

Address ______________________

______________________ Apt ______

City ______________________ State ______

Zip ____________ Phone (   ) ____________

Guaranteed. Period.®

LANDS' END DIRECT MERCHANTS

© 1992 SCIENTIFIC AMERICAN, INC

the world through repeated empirical observations. Popper agreed with the positivists that a theory can be true, but he rejected their contention that we can ever *know* it is true. "We must distinguish between truth," he says, "which is objective and absolute, and certainty, which is subjective."

Moreover, according to Popper, science cannot be reduced to a formal, logical system or method, as the positivists suggested. A theory is an invention, an act of creation, based more on a scientist's intuition than on preexisting empirical data. "The history of science is everywhere speculative," Popper says. "It is a marvelous history. It makes you proud to be a human being."

If science is a truly creative enterprise, however, then the world must unfold in fundamentally unpredictable ways. Recognizing this fact, Popper has waged a lifelong battle against determinism. "Determinism means that if you have sufficient knowledge of chemistry and physics you can predict what Mozart will write tomorrow," he says. "Now this is a ridiculous hypothesis."

Popper realized early on that quantum mechanics undercuts determinism by replacing classical certainty with "propensities." But even classical systems are inherently unpredictable, a fact that Popper claims to have discovered long before modern chaos theorists. In 1950 he found a theorem by a 19th-century French mathematician showing that an infinite number of geodesics, or shortest paths, can connect two points on a two-dimensional surface called a horned plane. This theorem demonstrates that "the world is chaotic," Popper says.

Popper first elucidated his ideas in 1934 in what is still his best-known book, *The Logic of Scientific Discovery*. It was so well received that in 1936 Popper, who was then teaching high school, was offered a philosophy professorship at Canterbury University College in New Zealand. After riding out World War II in the antipodes, Popper took a position at the London School of Economics and Political Science, where he remains a professor emeritus.

Popper's ideas, more than those of any other philosopher, have been warmly received by scientists, physicists in particular. The British physicist Hermann Bondi recently praised Popper's philosophy in *Nature* as "still the touchstone of whether one's ideas are scientifically meaningful." The admiration is mutual. Popper keeps reminding me that he was personally acquainted with such giants as Einstein, Schrödinger and Bohr. "I knew him quite well," Popper says of Bohr. "He was a marvelous phys-

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

icist, one of the greatest of all time, but a miserable philosopher."

Popper has generally been at odds with his fellow philosophers, whatever their outlook. His book *The Open Society and Its Enemies,* published in 1945, angered political philosophers of all stripes with its attacks on Plato, Marx, Hegel and others whom Popper accused of political dogmatism. Politics, even more than science, Popper insisted, requires the free play of ideas and criticism. Dogmatism inevitably leads not to utopia, as Marxists and fascists alike have claimed, but to totalitarian repression. (Some students of Popper, noting his own authoritarian demeanor, referred to his book as *The Open Society by One of Its Enemies.*)

Popper also wrangled with those

"Most philosophers are really deeply depressed because they can't produce anything worthwhile."

postwar philosophers who argued that philosophy is nothing but tautological word games. In his autobiography, *Unended Quest,* Popper recalls his legendary duel with a progenitor of this school, Ludwig Wittgenstein. Popper was giving a lecture at Cambridge in 1946, when Wittgenstein interrupted to proclaim the "nonexistence of philosophical problems." Popper demurred, saying there were many such problems, such as establishing a basis for moral rules. Wittgenstein, who was nervously playing with a fireplace poker, thrust it at Popper, demanding, "Give an example of a moral rule!" When Popper replied, "Not to threaten visiting lecturers with pokers," Wittgenstein stormed out of the room.

Popper is repelled by the currently popular view that science is driven more by politics and social custom than by a rational pursuit of the truth. He blames this attitude on a plot by social scientists to overthrow the traditional pecking order in science, which accords physicists the most legitimacy and social scientists the least. "This is a horrible affair," Popper growls, "and it will go on probably and become more and more horrible."

Yet scientists, too, can fall short of Popper's ideal. The growing subsidization of science by the public since World War II has led to "a certain corruption," he says. "Scientists are not as critical as they should be. There is a certain wish that people like you"—he jabs a finger toward my chest—"should bring them

© 1992 SCIENTIFIC AMERICAN, INC

Your Passport To Intellectual Adventure

The premier edition of the *Vintage* SCIENTIFIC AMERICAN *Wall Calendar* features reproductions of original engravings from the classic 19th century editions of SCIENTIFIC AMERICAN. Each engraving is accompanied by an excerpt from the original article's text.

The calendar is printed on replica parchment, measures 13" x 22" open, and has ample room for daily notes. Makes an impressive gift. Just $9.95 each ($1.50 S&H, $2.50 outside North America).

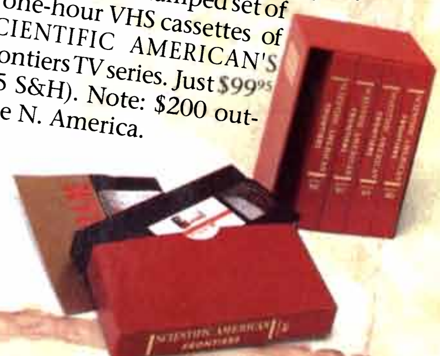
Fascinating insight and startling new discoveries await you in the 1992 SCIENTIFIC AMERICAN single-topic issue—*MIND AND BRAIN*. You'll journey deep inside the human mind for an in-depth look at how individual thoughts are related to one another; how memories are stored and much, much more. Plus, you'll examine breakthroughs in imaging technology, molecular genetics and psychopharmacology. Just $4.95 each ($1 S&H).

RUFUS·PORTER ENTERPRISES
A Scientific American Company

On Thursday, August 28, 1845, a new magazine made its first appearance. The magazine was called SCIENTIFIC AMERICAN and was the brainchild of Rufus Porter. In addition to being SCIENTIFIC AMERICAN'S founder and first editor, Mr. Porter was an inventor, renowned artist and champion of technological ingenuity for all of his 93 years. We at Rufus Porter Enterprises are dedicated to Mr. Porter's spirit and vision. The merchandise you see offered here is a combination of artistry and scientific achievement we believe he would applaud. If any product you ever order from us fails to meet or exceed your highest expectations, we insist you send it back for a courteous refund. Your complete satisfaction is guaranteed.

These 100% silk double width ties were designed in honor of Dr. Richard Leakey. Each is emblazoned with major Leakey finds of *Homo erectus, Homo habilis,* and *Australopithecus boisei* skulls. A portion of the proceeds helps fund Leakey work in Kenya. Gift-boxed in your choice of Red or Navy Blue $29.95 ea. ($2 S&H per order, $5 S&H outside North America). SAVE $4.95! Second tie only $25.

Peek behind the scenes at the Moscow Circus' engineering department. Discover the physics of roller coasters. Meet the Alaskan ground squirrel that drops its temperature to below freezing when hibernating. It's all here (and much more) on this handsomely packaged and gold stamped set of 5 one-hour VHS cassettes of SCIENTIFIC AMERICAN'S Frontiers TV series. Just $99.95 ($5 S&H). Note: $200 outside N. America.



Yes! Rush me the items I've selected!

Name ______________________

Address ______________________ Apt. ______

City ______________ State ______ Zip ______

Method of Payment (Please, No Cash!)

❑ Check ❑ Money Order ❑ Visa ❑ MasterCard

Card Number ______________________ Exp. Date ______

Signature ______________________

Item	Qty.	Price Ea.	Total	S&H Ea.	Total
Mind + Brain Special Issue		$ 4.95		$1	
Leakey Tie (Red)		$29.95		$2	
Leakey Tie (Navy Blue)		$29.95		$2	
Second tie just $25. Circle color desired: Red Navy Blue					
Frontiers 5-Tape VHS Set		$99.95		$5	
1993 Vintage Wall Calendar		$ 9.95		$1.50	

Rufus Porter Enterprises
SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue
New York, NY 10017-1111

Total Item Price
Total Sales Tax*
S&H Total
Grand Total

*Residents of CA, IA, IL, MA, MI, NY please add appropriate sales tax

11/92 RP 112A

Call 1-800-777-0444 To Embark On Your Adventure

© 1992 SCIENTIFIC AMERICAN, INC

before the public." He stares at me a moment, then reminds me that he did not seek this interview. "Far from it," he says. "You were trying to get me."

Popper is particularly dismayed at the way modern cosmologists cling to the big bang theory and "trample down" alternatives. He was thrilled by the expansionary universe model when it was first proposed in its simplest form more than a half century ago. "But then came the difficulties, and lots of additional hypotheses were needed," he says. Today, he adds, "the critical attitude in this field is very rare."

Popper has in the past irritated biologists by suggesting that Darwin's theory of evolution is untestable and even tautological (survivors survive) and hence not scientific at all. Popper is clearly reluctant to rehash this issue now, however. Although I ask him several times about his current position, he says only that he may have gone "a little too far" in his criticism.

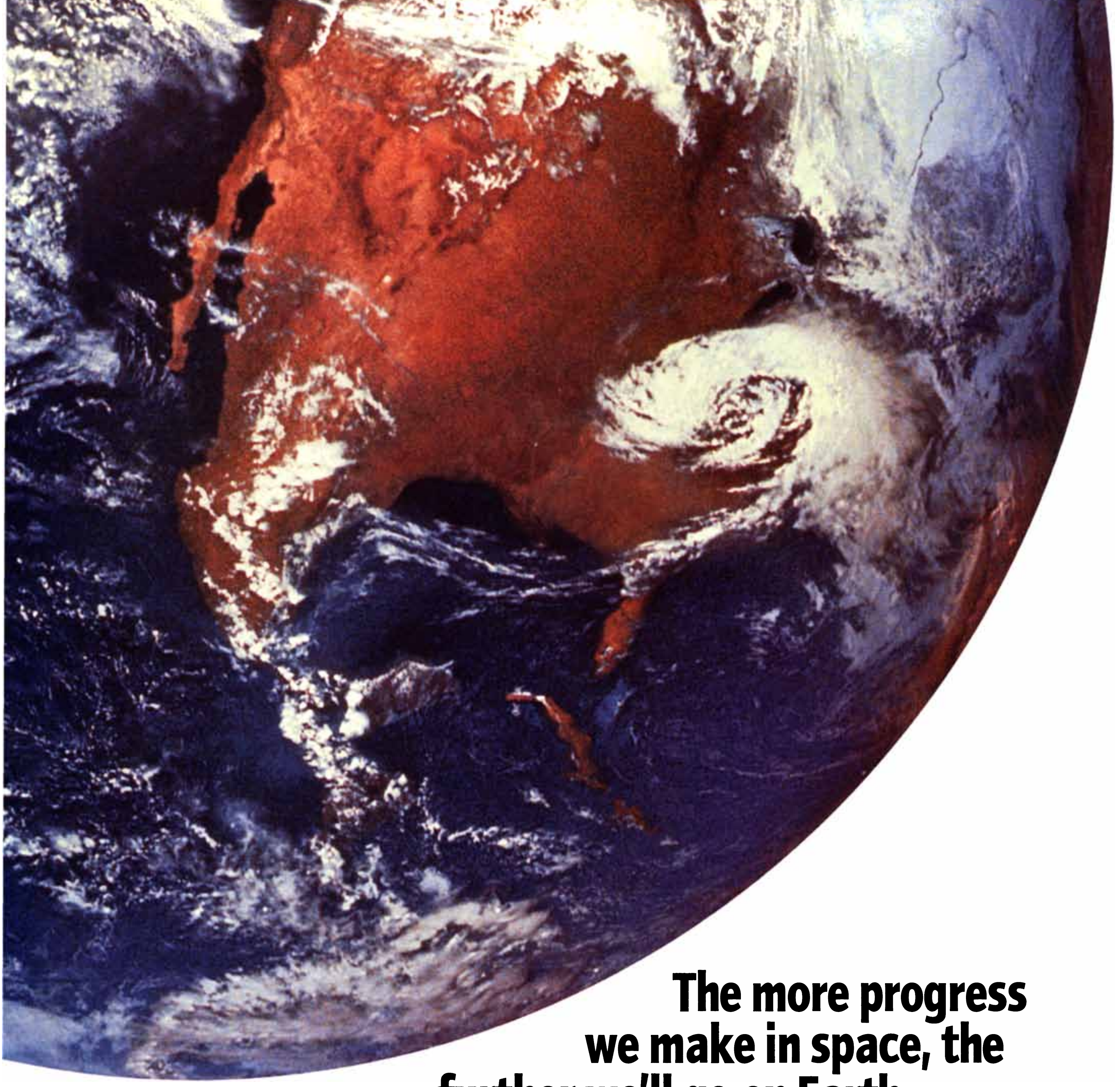
Rather than discussing Darwinism, Popper keeps returning to an unusual theory of the origin of life advanced by Günter Wächtershäuser, a patent attorney living in Munich who holds a doctorate in chemistry. Wächtershäuser proposes that life began not with the replication of biochemical units such as RNA, as many modern biologists believe, but with a biochemical process similar to the Krebs energy cycle in cells. ("I knew Krebs," Popper remarks.)

Popper feels some pride of ownership toward the new theory; Wächtershäuser developed it with Popper's encouragement after taking a summer school course from Popper a decade ago. "He had given up science, and I brought him back to his senses," Popper says. But he emphasizes that neither he nor Wächtershäuser is wedded to the proposal. "For God's sake, why shouldn't there be a third alternative?"

Popper's assistant pokes her head in the door to inquire, a bit grumpily, how much longer we intend to go on; we have already been talking for nearly three hours. I say I only have one more question: Why, in his autobiography, did Popper call himself "the happiest philosopher I know"? Popper grins wickedly. "Most philosophers are really deeply depressed because they can't produce anything worthwhile," he replies.

Noting his assistant's look of alarm, Popper's smile fades. "Perhaps you should not write this," he says to me. "I have enough enemies." He thinks a moment, then continues, his eyes brightening again: "I feel in all modesty overwhelmed with ideas, of which I can only put some down, and that's wonderful." —*John Horgan*

© 1992 SCIENTIFIC AMERICAN, INC



A view of the United States as seen from space.
Photo courtesy of NASA.

The more progress we make in space, the further we'll go on Earth.

America needs space to grow. Advances in science, medicine and technology will be enhanced, even markedly increased, through our efforts in space. Ensuring America's preeminence in space is a national priority. As one of NASA's top contractors, Lockheed is helping to lead this important effort.

We're at the forefront of space technology today. Our expertise extends from engineering and science support of Johnson Space Center to leadership in vital programs like Space Station *Freedom*, Space Shuttle Processing and the Hubble Space Telescope.

National priorities such as the space program demand advanced technology, reliable systems and affordable solutions. They demand a premier aerospace company. They demand Lockheed.

Lockheed leads.

Lockheed

Watch NOVA on PBS, Tuesdays at 8 p.m.

© 1992 SCIENTIFIC AMERICAN, INC

SCIENTIFIC AMERICAN

Health Care Reform

Medical costs are rising rapidly, and millions of people have no health care coverage. The nation urgently needs a universal insurance program

by Rashi Fein

The U.S. faces two challenging health care problems. The first is that an estimated 35 million Americans have no medical insurance and millions more have only limited coverage. This situation is compounded by the fact that most Americans believe that medical care is a right: people should not be denied necessary treatment because of their income. Despite disagreements about particulars—such as what constitutes necessary or adequate treatment and how much services should cost—there is general agreement that the U.S. has failed to develop a system for the equitable distribution of health care.

The second, newer problem involves the effect of rising health-related costs on the nation's long-run economic prospects. In 1940 health care absorbed $4 billion, a mere 4 percent of our gross national product (GNP). In 1990 such expenditures equaled $666 billion, or 12.2 percent of the GNP; projections suggest that in 1992 the country will spend more than $800 billion on medical care, or 13.4 percent of the GNP.

RASHI FEIN is professor of the economics of medicine at Harvard Medical School. Although his work has included studies of manpower supply and the financing of medical education, his current work is directed at health care reform. Fein, who received his doctorate from Johns Hopkins University, served on President John F. Kennedy's Council of Economic Advisers. He is the author of several books on the economics of health care and is a member of the Institute of Medicine.

The increasing proportion of the GNP spent on health is associated with a declining share of the GNP spent on other needs such as education and repair of the infrastructure, as well as on research and development. In 1990 businesses spent 61 percent of pre-tax profits and 108 percent of after-tax profits on health care benefits for employees (as opposed to 20 and 36 percent, respectively, in 1970). Health payments were 15.3 percent of total federal expenditures; 11.4 percent of state and local budgets went to health. These allocations reduce the funds available for meeting other government commitments and for investing in economic opportunities that contribute to long-term growth.

The U.S. also spends more of its gross domestic product (GDP)—the value of items produced solely within U.S. borders—on health care than any of the 23 other members of the Organization for Economic Cooperation and Development (OECD). In 1989, for example, the U.S. spent 11.8 percent of its GDP on health. In contrast, Canada spent 8.7 percent; 16 other nations spent less than 8 percent. The private sector financed about 60 percent of the U.S. expenditures, as opposed to only 20 percent in all OECD countries combined. As a result, American companies have proportionately less capital to invest, thereby jeopardizing the country's international competitive position.

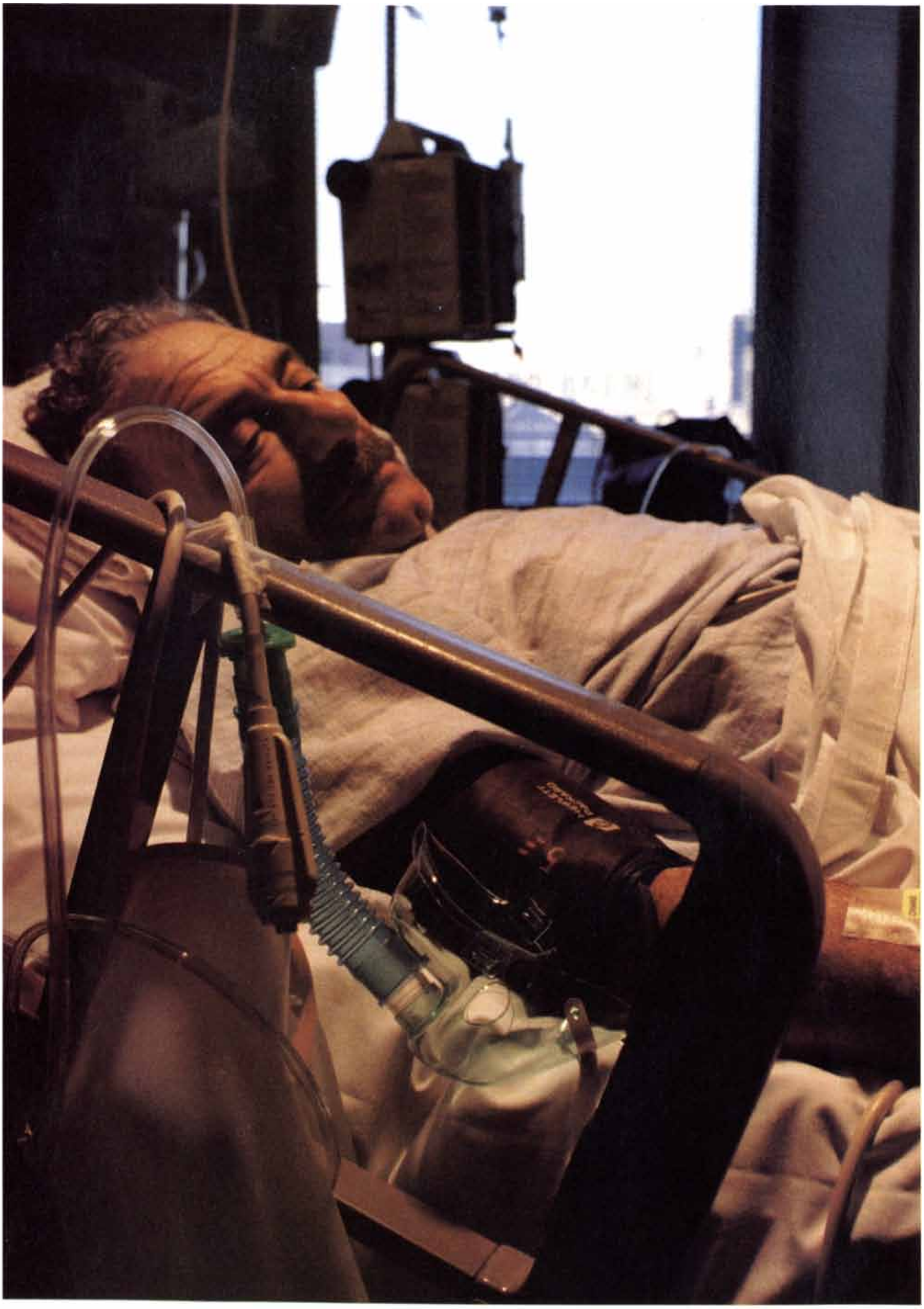
Clearly, the development of a novel health care system is ethically and economically imperative. Although building on the current foundation has a certain political appeal, that basis is severely flawed. The country needs to erect a new structure, one that provides universal, or national, health insurance and one that helps to contain health care expenditures.

Discussions about national health insurance in the U.S. began more than 75 years ago. Prompted by the Progressive Movement and building on successful state initiatives on workmen's compensation, state legislatures began to discuss such plans. Over time, visions of the goals, structures, administration and funding for such a program have inevitably changed. *A Discursive Dictionary of Health Care,* published in 1976 by Congress, hints at the ever changing perceptions of health care reform. The definition given for national health insurance is "a term not yet defined in the U.S." As the divergent bills now before Congress demonstrate, this term remains undefined.

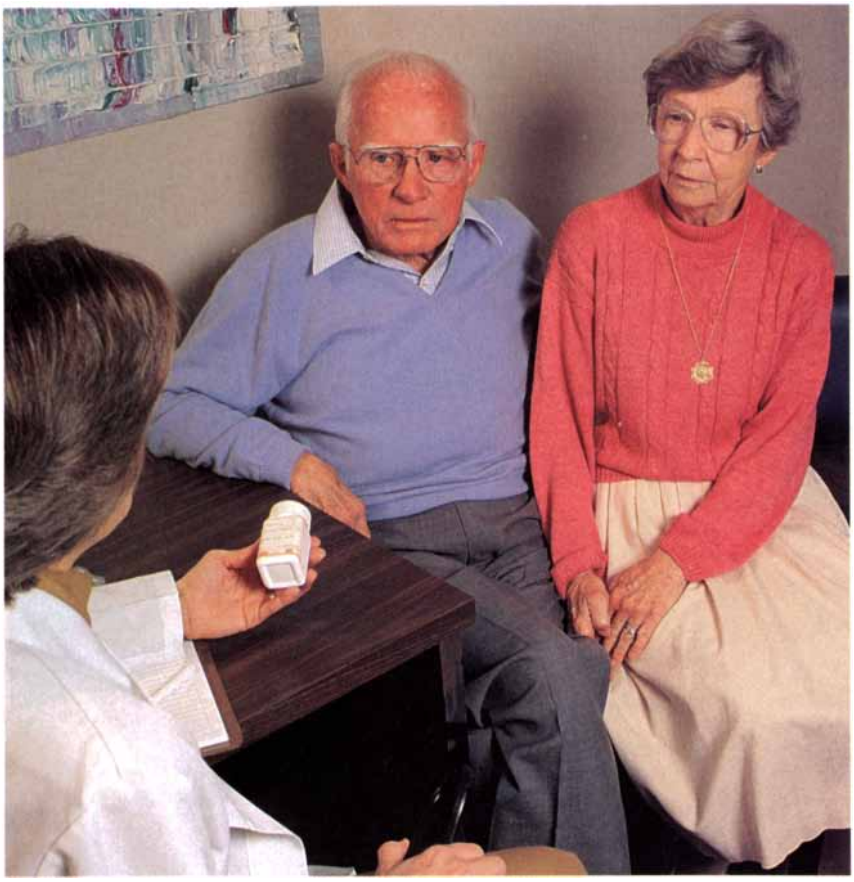
Yet the major issue has remained fairly consistent. The first arguments for national health insurance were framed in terms of equity: remove financial barriers for people whose medical care costs were more than they could afford and who had inadequate coverage or no insurance at all. The source of these gaps in insurance has been the continuing inability of private (as opposed to governmental) insurance to provide affordable coverage equitably.

RISING MEDICAL COSTS threaten people who lack medical insurance—a group numbering more than 35 million in the U.S. To prevent a social and economic crisis, sweeping reforms in the current insurance system must be instituted.

© 1992 SCIENTIFIC AMERICAN, INC



© 1992 SCIENTIFIC AMERICAN, INC



COVERAGE FOR THE ELDERLY and the disabled is guaranteed by Medicare. Before the creation of that program in 1965, many of the elderly were not covered by insurance or had to pay exorbitant premiums because they were considered bad risks. Medicare, which does not discriminate against persons at high risk, could serve as the model for a universal health insurance program.

The insurance gap stems from the way private insurance changed during and after World War II. During the war, the government froze wages, and management and labor were encouraged to bargain over medical coverage. Employees desired insurance because of the greater ability of medicine to cure illness; employers offered this fringe benefit because it was, at least at first, relatively inexpensive. These conditions led to growth for the insurance industry, a pattern that continued after the war, when the National Labor Relations Board ruled that employers unwilling to bargain over health care benefits were guilty of an unfair labor practice. Consequently, employment-based private health insurance grew rapidly during the 1940s and 1950s.

Employer-based private insurance, however, fails to provide coverage for those who are not employed or who work temporarily or sporadically. It also misses people whose employers, for reasons including the high costs of health insurance, do not provide insurance benefits. Another problem emerged as well. Initially, premiums were based on a community-rating principle: that is, payments reflected the average cost of care for all subscribers. The amount paid by a group or individual did not depend on age, gender, health status or the previous or anticipated use of health services. This strategy implied a subsidy from younger, healthier members of the population to its older, sicker members. Buyers considered the subsidy, a form of risk averaging, one of the desired attributes of insurance.

But community rating did not survive for long in a competitive world. Private insurance companies sought new markets by basing premiums on the health care costs incurred by specific subgroups of subscribers. This approach became known as experience rating, and it encouraged insurers to look for lower-risk individuals or groups (including employees of particular firms) and to offer them reduced rates. This tactic had two unfortunate consequences: it became advantageous for persons or groups with lower risks to dissociate from persons with higher risks, and groups with people at higher risk, such as the elderly, found that their premiums increased.

Cost sharing therefore took place solely within particular groups or subgroups—but not across groups. Inevitably, as those with less risk moved away from those with higher risk, the average risk for those still in the community group increased, as did their premium costs. In turn, individuals who perceived themselves as having lower than average risk left the community, causing a further escalation of costs for those who stayed. A vicious circle was established.

The detrimental ramifications of experience rating are apparent today. Because the costs of health care have exploded, the difference in premiums between high- and low-risk individuals and between community and experience rating has increased. These escalating costs have variably affected businesses as well as individuals. In particular, firms, such as the Ford Motor Company, that have older employees who use more health services, are at a competitive disadvantage. Companies such as Honda, whose employees tend to be younger and healthier, do not incur high costs. (This drawback is, of course, made even more severe by the fact that older firms provide coverage for early retirees, a category that does not exist in the younger firm.)

Experience rating has also drawn attention to preexisting conditions—health problems stemming from a condition that an individual has before he or she is covered by a policy. This issue is likely to become more pervasive as genetic research enables more precise measurement of individual risk factors: one's genes will become the ultimate preexisting condition. Unless the country changes its current health insurance system, individuals at above-average risk will find it ever more difficult to obtain employment that includes sufficiently comprehensive coverage.

Many factors help to explain the explosion in medical costs that has increased the vulnerability of the American economy and exposed the weaknesses found in the experience-rating system. The population has grown (and it continues to do so). Elderly people, whose health care costs are high, represent a burgeoning proportion of the population. The economy has suffered from general inflation. In addition, because health care is a service industry, it is a sector of the economy with low productivity gains and substantially large increases in costs.

And after all, the base cost of provid-

© 1992 SCIENTIFIC AMERICAN, INC

ing medical care has grown. The threat of malpractice suits has increased the pressure to practice defensive medicine. Physicians routinely perform tests and procedures that have questionable medical value but considerable legal consequences. Costs and expenditures have also risen for the best of reasons: medicine's improved ability to help people.

Furthermore, the health care community has never operated within any budgetary constraints. Many patients are sheltered from costs because their insurance covers them. Physicians, aware that their patients are protected, have been free to set fees in an unregulated market. Indeed, U.S. physicians earn substantially more than doctors in any other OECD country.

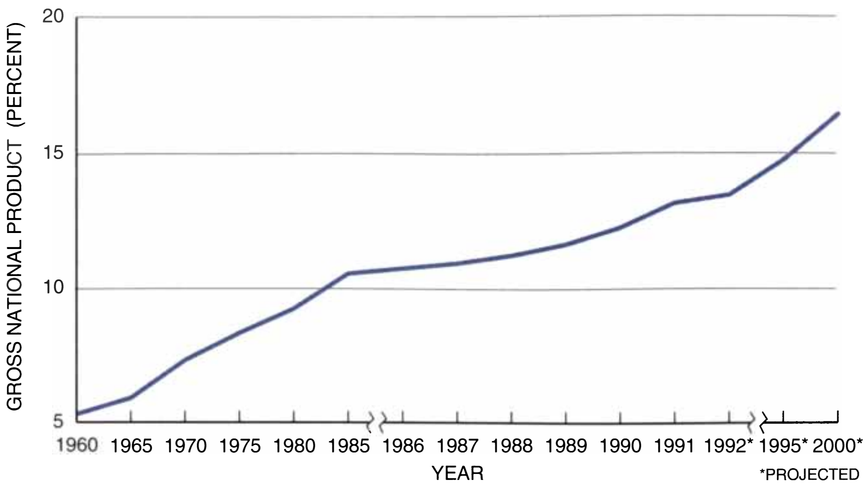
Escalating medical costs have also undermined the employer-based private insurance structure that emerged after World War II. The high cost of coverage has encouraged employers to limit their expenditures. They have made cutbacks in benefits, increased employee cost sharing and reduced coverage for dependents.

It has become obvious that employer-based private insurance operating within a system dominated by experience rating cannot reach everyone. There are now, as there were in the 1950s, three groups likely to be underinsured or inadequately insured: unemployed people, those whose employers do not provide insurance and applicants with preexisting conditions. Access to insurance has become an especially significant issue as medicine's ability to treat and prevent illnesses successfully has improved and as the costs of such lifesaving interventions have risen.

It is possible to erect a universal insurance program that would avoid the gaps and adverse effects of the current system. The concept is straightforward: every person would be enrolled in the same financing program, one similar to the Medicare model. Under the Medicare program for the aged, disabled and people with end-stage renal disease, beneficiaries seek care from diverse sources even though the same insurance program covers them all.

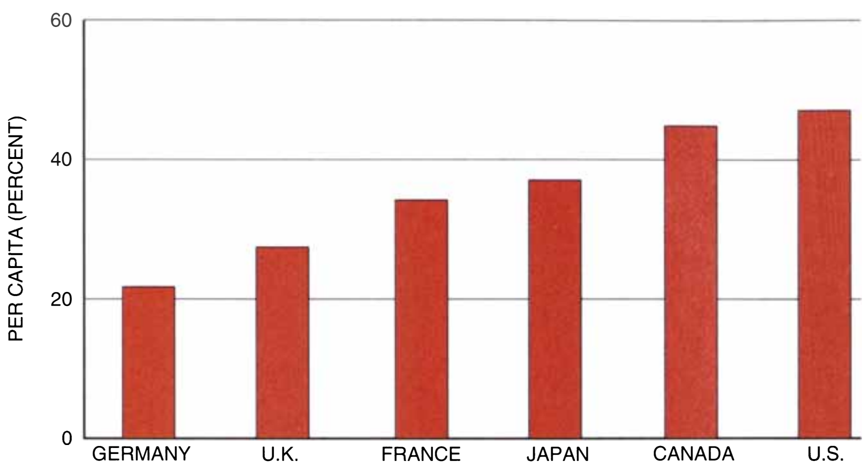
Payments to Medicare are not related to an individual's health status or to current or projected future use of health services. Nor would it be necessary to make them so in an expanded system that enrolled all U.S. citizens. Experience rating disappears—and with it the rationale for discriminating against high-risk users. The impetus to shift costs from one program

U.S. Health Care Costs



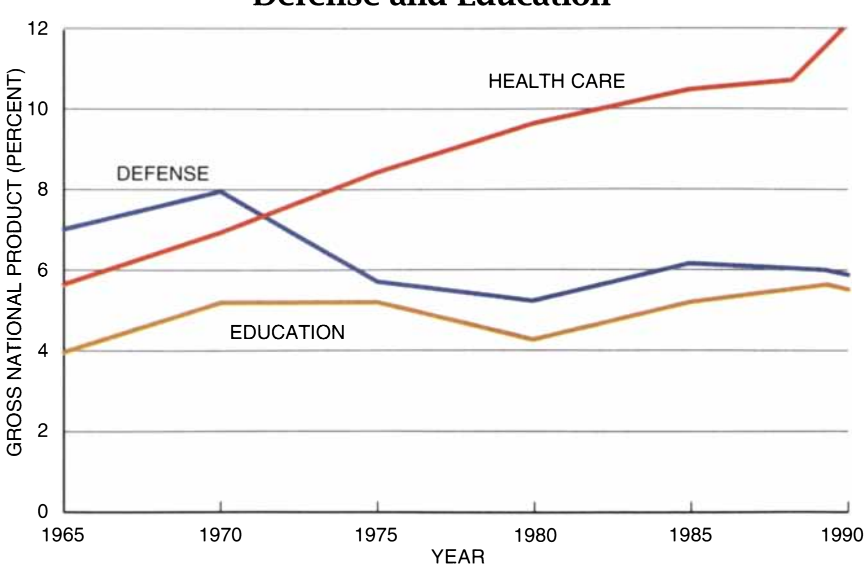
SOURCE: Health Care Financing Administration

Increase in Health Care Spending, 1980–1989



SOURCE: Organization for Economic Cooperation and Development

Comparison of Costs for Health Care, Defense and Education



SOURCE: Consumer Reports, *U.S. National Center for Education Statistics, Health Care Financing Administration, U.S. Office of Management and Budget*

© 1992 SCIENTIFIC AMERICAN, INC

to another falls by the wayside when there is only one program.

Moreover, a single universal program severs the link to employment. In contrast, pay-or-play proposals, often presented as a pragmatic compromise between a radically new system and current practices, expand that linkage. Such policies would require employers to either "pay" (forward a specified amount to a government fund that would be used to provide insurance) or "play" (provide their employees with health insurance that meets a predetermined benefit standard).

But because pay-or-play proposals build on the existing model of employment-based insurance, they would continue patterns of inefficiency. They would have to deal with coverage for part-time work and self-employment. They would have to have separate mechanisms to provide continuity of coverage for individuals who change employers or location. They would have to account for the differing ability of employers to finance insurance and provide appropriate subsidies for "marginal" firms as well as for start-up enterprises. Pay-or-play policies would also have to avoid propagating incentives to hire individuals with lower risks whose premiums would be less costly.

Although it is possible to design special provisions to meet these dilemmas, each attempt to compensate for the complexities of an employment-based approach comes at a price. The system becomes more cumbersome and suffers from increased administrative duties and higher administrative costs.

Pay-or-play proposals would have to be carefully tuned and retuned in order to maintain the desired balance between the private and the public provision of insurance. A relatively low pay provision (in other words, too low a tax) would make the play alternative unattractive and would lead to the decline of private insurance. Conversely, too high a pay provision would force some employers not offering insurance to play—even though that choice might place an onerous burden on them, in the form of lower profits, or on their employees, in the form of lower wages or unemployment.

If everyone is enrolled in the same program, however, a number of otherwise complex financing issues are automatically resolved. Again, Medicare provides a helpful precedent. Medicare is financed through a combination of sources. In 1992 an estimated 58 percent of total income to the fund will be derived from payroll taxes, 25 percent from general revenues and 9 percent from individual premiums. Just as none of these payments are related to the individual's health status, the same would be true in an expanded universal program that included all Americans. Experience rating and its unfavorable effects would disappear.

A universal program could rely on any of the financing sources used in Medicare, or it might be financed by additional revenues such as those from a value-added tax. The choice of financing sources is vitally important, and the initial political and economic difficulties brought about by shifting a large proportion of the estimated $460 billion of private health care funding into the public sector are considerable. Yet these are surmountable, one-time concerns and are eased by the recognition that on average the increase in taxes to finance universal coverage would be balanced by the decline in private payments for insurance premiums and care. The program would require a shift, not an augmentation, in health care financing.

A single universal program has another positive feature—one that becomes apparent when we compare Social Security with welfare and Medicare with Medicaid. Social Security and Medicare do not discriminate on the basis of income. Conversely, welfare and Medicaid are solely for the poor. The first two are strong programs that, by including everyone, protect low-income individuals from the vagaries of funding; the fates of diverse income groups are inexorably intertwined.

In contrast, we fund welfare and Medicaid for "them"—and we all know or think we know who "they" are. These programs rely on the milk of human kindness, but this milk sometimes curdles. To minimize the development of disparities in access and in quality, it is important to erect programs that do not enroll persons according to socioeconomic characteristics and sources of funding.

Yet this is the risk we run with strategies, such as pay-or-play plans, that combine private and public approaches. The publicly funded programs would have a disproportionate number of low-income or high-risk people for whom employers prefer to pay. Their per capita medical costs would be higher because medical needs reflect health status, which is, in turn, affected by housing, education and employment.

Once again, government plans would appear to be more expensive than private initiatives, presumably because government is less efficient, thereby providing a rationale for cutting public budgets and for underfunding. Whatever level the publicly funded initiative begins at, over time a program for the poor will become a poor program.

Enrolling everyone in the same plan will certainly not solve or even address all the shortcomings of the current system. But the approach would enable us to tackle the issue of cost containment in a direct way.

A single enrollment program means a single payer or, more appropriately, a single purchaser of services. This situation leads to the standardization of forms, to electronic billing and to various measures that reduce confusion, delay and the costs of administration in today's system. The purchaser would not only pay the bill for services but also accept a broader responsibility to focus on issues of quality of care, unnecessary services and value for money. Each of these activities contributes to cost control.

In addition, a universal program that raises funds through taxes and premium payments cannot survive for long if it neglects cost considerations. If it cannot fund its obligations, it will be forced to increase its revenues or shift costs to patients—and it will fail. Because the single insurance approach cannot shift costs to another insurer, it must achieve its goal within a budget.

Therefore, the plan would have to set an overall budget for health care as well as similar budgets for physician and hospital services. As demonstrated by many countries, such as France and Japan, negotiations between all affected parties, including government and health care providers, can give rise to a budget. This political process weighs the costs of health care against alternative public programs and private-sector expenditures.

Measures such as prospective budgets for institutional care, capitated (non-fee) payments for people enrolled in Health Maintenance Organizations, and other approaches to managed care would assist in meeting budgetary constraints. There is no intellectual challenge to developing the tools necessary to do the task.

Great difficulty may arise, as it does with all large government programs, in choosing a budget that reflects the public view of competing consumption and investment opportunities (such as education). In the 1960s and early 1970s voters feared that the government was profligate and that national health insurance would bankrupt the nation—especially a nation that associated larger expenditures with higher quality. Today the fear is the reverse, that legislators will be parsimonious and that the nation will not spend as much as is essential to maintain the medical care infrastructure.

© 1992 SCIENTIFIC AMERICAN, INC

Of course, there can be no guarantee that underspending will not occur. Nevertheless, it is possible to structure national health insurance to reduce the likelihood of such a phenomenon. The most important protective device is universality—the fact that everyone is covered by the same financing scheme. Although individuals will be able to opt out of the publicly funded program in favor of private insurance, the incentive to exercise that option can be reduced by requiring payment of the universal tax even if one elects to receive privately funded care. Moreover, stipulating that providers be entirely in or out of the program would guard against underspending.

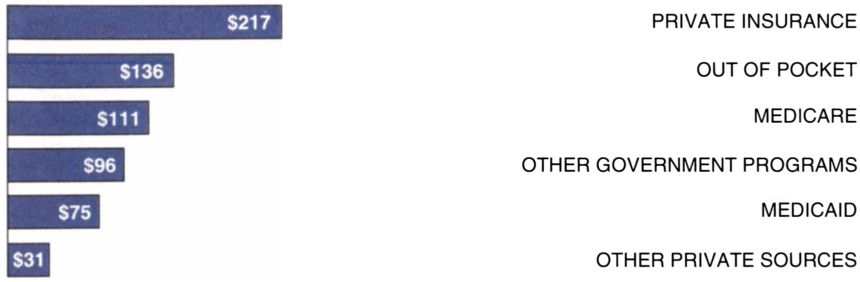
To formulate workable health care budgets, policymakers must become more concerned about the supply of health resources. The proliferation of high-technology equipment and services and the growth of financially rewarding specialties—surgery, anesthesiology, radiology and subspecialties of internal medicine—have contributed to expanding costs. If the system does not solve the monetary and other incentives that have led to a declining interest in family medicine, general medicine and pediatrics, costs will not be contained nor quality sustained.

Not even the best national health insurance program can succeed in assuring access to care at a reasonable, responsible expenditure level if health restructuring is limited to insurance enrollment, financing and payment. Medical education and training must be reformed. So also must the financing of medical education. Its high cost can lead to indebtedness and consequently cause students to choose lucrative specialties.

Many academic health analysts believe a remodeled U.S. health care system would be able to deliver services to all Americans for the $800 billion now spent on care for only some Americans. These experts say the system could do so without denying anyone care that is medically necessary. They point to the large savings that would accrue if unneeded care were eliminated and patterns of administration streamlined. Indeed, some estimates suggest that as much as 25 percent of performed procedures are not required.

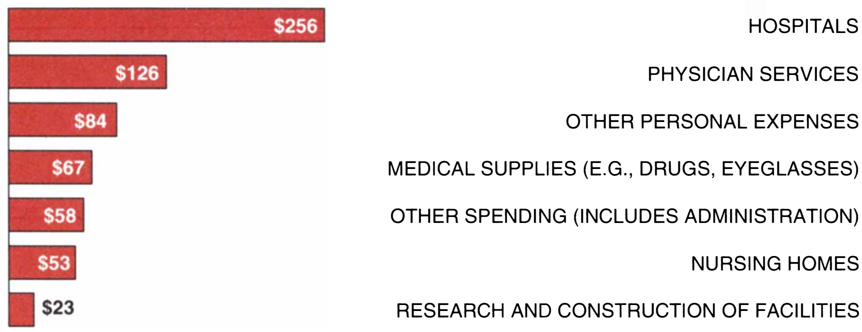
Nevertheless, the savings would occur only over a number of years. Determining what constitutes appropriate care would require research into and consensus about the meaning of unnecessary care. Patterns of practice would have to be altered. We would have to pursue an

How Health Care Is Financed, in Billions



SOURCE: Health Care Financing Administration, 1990

Where Health Care Funds Are Spent, in Billions



SOURCE: Health Care Financing Administration, 1990

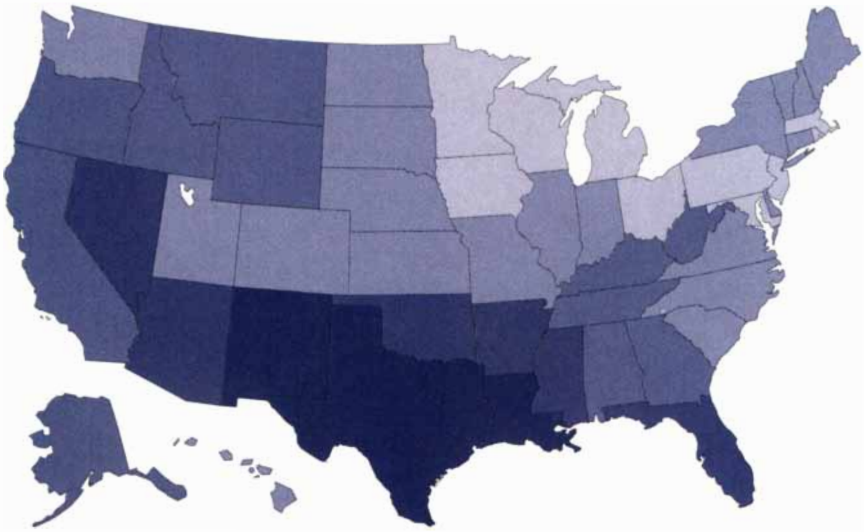
Comparison of Physicians' Incomes

	AVERAGE PHYSICIAN'S INCOME IN 1986 U.S. DOLLARS	RATIO OF PHYSICIAN'S INCOME TO AVERAGE INCOME	PHYSICIANS PER 1,000 POPULATION, 1981–1983
AUSTRALIA	34,191	2.26	1.9
CANADA	70,144	3.47	1.9
DENMARK	39,061	2.01	2.4
FINLAND	35,558	1.82	2.2
FRANCE	NOT AVAILABLE	3.27 (1979)	2.2
WEST GERMANY	91,244	4.28	2.4
IRELAND	17,830	1.08	1.2
ITALY	NOT AVAILABLE	1.10 (1981)	1.3
JAPAN	56,437	2.46	1.4
NORWAY	31,664	1.38	2.0
SWEDEN	NOT AVAILABLE	1.80 (1983)	2.4
SWITZERLAND	118,501	4.10	1.6
U.K.	33,615	2.39	1.3
U.S.	119,500	5.12	1.9

SOURCE: Organization for Economic Cooperation and Development, 1986

© 1992 SCIENTIFIC AMERICAN, INC

Percentage of State Populations That Lack Health Insurance



BELOW 10 PERCENT
10 TO 14 PERCENT
15 TO 19 PERCENT
(INCLUDES THE DISTRICT OF COLUMBIA)
20 TO 24 PERCENT
25 TO 26 PERCENT

SOURCE: General Accounting Office, 1988

Characteristics of the Insured and Uninsured

	EMPLOYMENT-RELATED COVERAGE	OTHER PRIVATE COVERAGE	PUBLIC COVERAGE ONLY	UNINSURED
EMPLOYMENT				
FULL TIME	81.9	2.6	2.8	12.7
PART TIME	59.7	8.8	7.5	24.1
SELF-EMPLOYED	54.7	20.6	1.8	22.9
TYPE OF INDUSTRY				
PERSONAL SERVICES	52.5	9.5	6.6	31.5
CONSTRUCTION	60.6	6.8	1.9	30.6
SALES	68.3	6.3	4.0	21.4
MANUFACTURING	85.9	2.8	1.0	10.3
SIZE OF ESTABLISHMENT				
FEWER THAN 10 WORKERS	56.5	13.1	4.2	26.3
MORE THAN 500 WORKERS	89.4	1.3	3.2	6.1
HOURLY WAGE				
$3.50 OR LESS	53.9	6.0	10.1	30.1
$3.51 TO $5.00	56.3	6.0	7.3	30.4
MORE THAN $15.00	91.3	1.9	1.7	5.1
ETHNIC GROUP				
WHITE	69.1	11.7	6.8	12.4
BLACK	48.5	4.5	25.1	22.0
HISPANIC	45.9	4.2	18.3	31.5
AGE				
UNDER 6	62.5	5.0	15.8	16.7
19 TO 24	55.2	8.1	6.5	30.2
65 AND OLDER	35.4	39.3	24.4	0.9

SOURCE: Department of Health and Human Services

All figures are percentages. Because of rounding, some totals do not equal 100 percent.

active labor-market policy to assure jobs for those who would be displaced by change.

But what if these savings were insufficient to permit the system to deliver, at reasonable expense, all of the services that providers might want to offer or that patients might seek? What if the capacity of the system proved temporarily insufficient?

I must emphasize that I do not believe that in a reorganized system as richly endowed as ours, the public would face queues and long waiting times. I do believe, however, that we should recognize the possible need for rationing. My dictionary defines the verb "to ration" as "to distribute equitably." In health care this means the use of nonincome criteria to guide distribution of services: the use of standards that give highest priority to medical needs, urgency of treatment and possible benefits.

Consensus on these matters is not easily arrived at, but solutions exist. In other nations the government has often resolved these issues by delegating the decision-making process to physicians who allocate and control information and resources, determine priorities and perform rationing.

Another strategy is for the public to assume more control and make rationing explicit. The state of Oregon engaged in just such a public debate to determine the specific services that would and that would not be provided under its Medicaid program. But there were a lot of problems. The proposed rationing scheme appeared rigid because it did not adequately take into account the physician's judgment about individual cases. It also disproportionally hurt poor mothers and children. The federal government did not approve the implementation of the program, arguing that it discriminated against the disabled.

The Oregon experience was, however, only the first attempt to impose rationing on part of the population at the state level. The effort illustrated the difficulty of achieving consensus but also demonstrated that through open debate the political process can construct and refine explicit rationing standards. In any event, a universal health insurance program does not create a need for rationing; rather it substitutes an equitable rationing scheme for the price-rationing system that now exists.

To be viable, a health care system must incorporate a sense of community, a sense of solidarity with others. That demand, it seems to me, argues in favor of a plan administered by states rather than by a distant federal govern-

© 1992 SCIENTIFIC AMERICAN, INC

ment. A series of insurance packages administered by the states, rather than one federal program, would not violate the concept of universality. The details of the meshing of federal and state responsibilities would be important. The federal government would have to define benefits and cost-containment goals, as well as provide some funding. But in my judgment, the gains associated with state administration and more direct citizen involvement would outweigh the additional complexity.

In discussing health care reform, other economists and health care experts emphasize the nature of the delivery system and the role of managed care, the role assigned to cost sharing and the type of negotiation process used to develop an appropriate fee structure. I do not deny the importance of these matters or suggest that the framework I have outlined cannot be altered in several respects without compromising the basic integrity of the proposals.

At the same time, I do mean to emphasize that we need to consider fundamental choices. We must choose between universal social insurance and the continuation of employment-based insurance, between community and experience rating and between cost containment and cost escalation.

The approach I propose has some disadvantages. In addition to being nonincremental, it increases the role of government. It also requires concern about the effects of restructuring, the fate of displaced workers and the consequences of shifting large sums from the private to the public sector. Yet a system analogous to an improved Medicare for all has the boon of simplicity. It can be explained and can be understood and is workable. The plan spells out the equity and expenditure issues efficiently.

HEALTH INSURANCE was readily provided by employers during the 1940s and 1950s. Medical care was at that time less expensive, and many people whose medical expenses had been paid by the federal government during World War II expected employers to provide similar coverage. This advertisement for insurance appeared in *Fortune* magazine in September 1945.

Citizens should not adopt the posture that many analysts and politicians have chosen: to reject what they agree is an efficient and workable program because they presume it is too big a step and cannot pass the legislative hurdles. The less efficient incremental approaches that seem "better" may prove as challenging to enact as the single enrollment program outlined here. Nothing will come easily, but on health insurance reform, as on other matters, the public may be ahead of the representatives. A Kaiser/Louis Harris poll at the end of June found that adults preferred a single-payer plan over a pay-or-play program by a margin of 35 to 29 percent—even though the latter has the endorsement of major congressional figures and leaders in the business community.

While the incremental pay-or-play program might be viewed as an "establishment" proposal that has received considerable attention and publicity, the more comprehensive single-payer proposal was preferred by 41 to 29 percent of Democrats and 38 to 29 percent of independents. Republicans did favor the pay-or-play approach, but only by 31 to 26 percent, despite the president's criticisms of the single-payer approach as socialistic and leading to long lines.

Election year is hardly the time to accept the pundits as the true experts on what is politically feasible. It is certainly not the moment to reject comprehensive reform on the grounds that it is not doable. Rather all of us must become more knowledgeable about and must debate each of the options that could improve the health of our society.

FURTHER READING

MEDICAL CARE, MEDICAL COSTS: THE SEARCH FOR A HEALTH INSURANCE POLICY. Rashi Fein. Harvard University Press, 1986.

BALANCING ACCESS, COSTS, AND POLITICS: THE AMERICAN CONTEXT FOR HEALTH SYSTEM REFORM. John Holahan, Marilyn Moon, W. Pete Welch and Stephen Zuckerman. Urban Institute Press, 1991.

SERIOUS AND UNSTABLE CONDITION: FINANCING AMERICA'S HEALTH CARE. Henry Aaron. Brookings Institution, 1991.

© 1992 SCIENTIFIC AMERICAN, INC

The Expansion Rate and Size of the Universe

The age, evolution and fate of the universe depend on just how fast it is expanding. By measuring the size of the universe using a variety of new techniques, astronomers have improved estimates of the expansion rate

by Wendy L. Freedman

Our Milky Way and all other galaxies are moving away from one another as a result of the big bang, the fiery birth of the universe. During the 20th century, cosmologists have discovered this expansion, detected the microwave background radiation from the original explosion, deduced the origin of chemical elements in the universe and mapped the large-scale structure and motion of galaxies. Despite these advances and many others, elementary questions remain unanswered. How long ago did the colossal expansion begin? Will the universe continue to expand forever, or will the universe eventually be halted by gravity and then collapse back on itself?

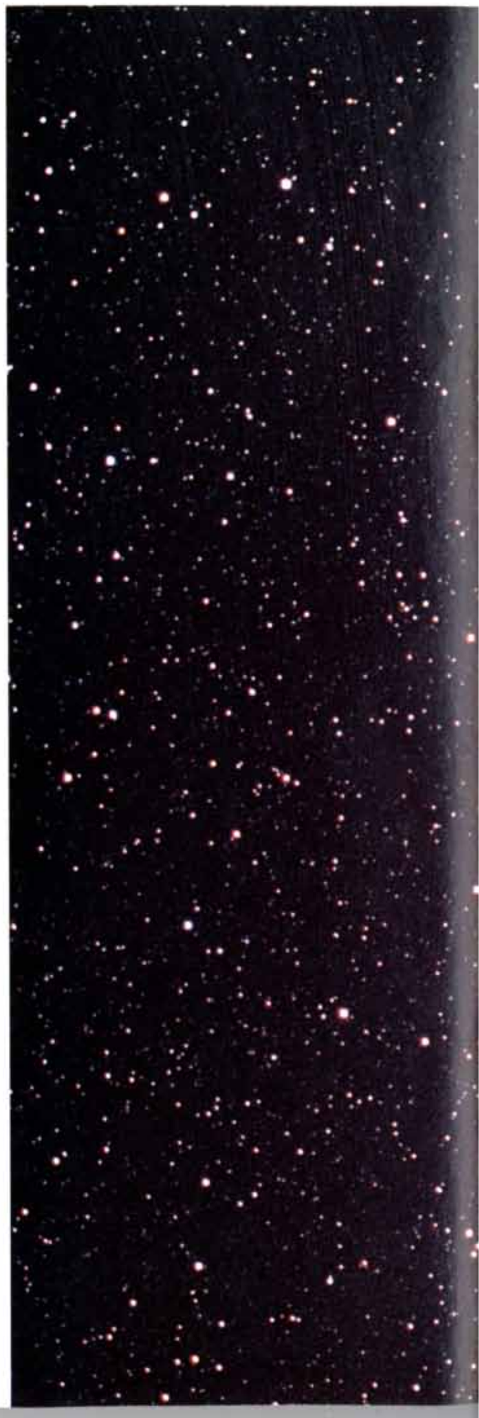
For decades cosmologists have been attempting to answer such questions by measuring the size scale and expansion rate of the universe. To accomplish this task, astronomers must determine both how fast galaxies are moving and how far away they are. Techniques for measuring the velocities of galaxies are well established, but estimating the distances to galaxies has proved to be a far more challenging task. During the past decade, several independent groups of astronomers have developed better methods for measuring the distances to galaxies, leading to completely new estimates of the expansion rate.

At present, several lines of evidence point toward a high expansion rate, implying that the universe is relatively young, perhaps only 10 billion years old. They also suggest that the expansion of the universe may continue indefinitely. Yet for many reasons, my colleagues and I do not consider the evidence definitive, and indeed we actively debate the merits of our techniques.

An accurate measurement of the expansion rate is essential not only for determining the age of the universe and its fate but also for constraining theories of cosmology and models of galaxy formation. Furthermore, it is important for estimating fundamental quantities ranging from the amount of nonluminous matter in galaxies to the size of clusters of galaxies. And because accurate measurements of distance are required for calculating the luminosity, mass and size of astronomical objects, the issue of the cosmological distance scale, or determination of the expansion rate, affects, to a greater or lesser extent, the entire field of extragalactic astronomy.

Astronomers began measuring the expansion rate of the universe some 60 years ago. In 1929 the eminent astronomer Edwin P. Hubble of the Carnegie Institution discovered that nearly all gal-

WENDY L. FREEDMAN is a member of the staff at the Carnegie Observatories in Pasadena, Calif. Born in Toronto, she received a Ph.D. in astronomy and astrophysics from the University of Toronto in 1984. She became a Carnegie postdoctoral fellow and, in 1987, was the first woman to join the staff. She is now a member of the Hubble Space Telescope Key Project Team for the Extragalactic Distance Scale. Her interest, outside of astronomy, is spending time with her family: husband Barry, daughter Rachael, 5, and son Daniel, 4.



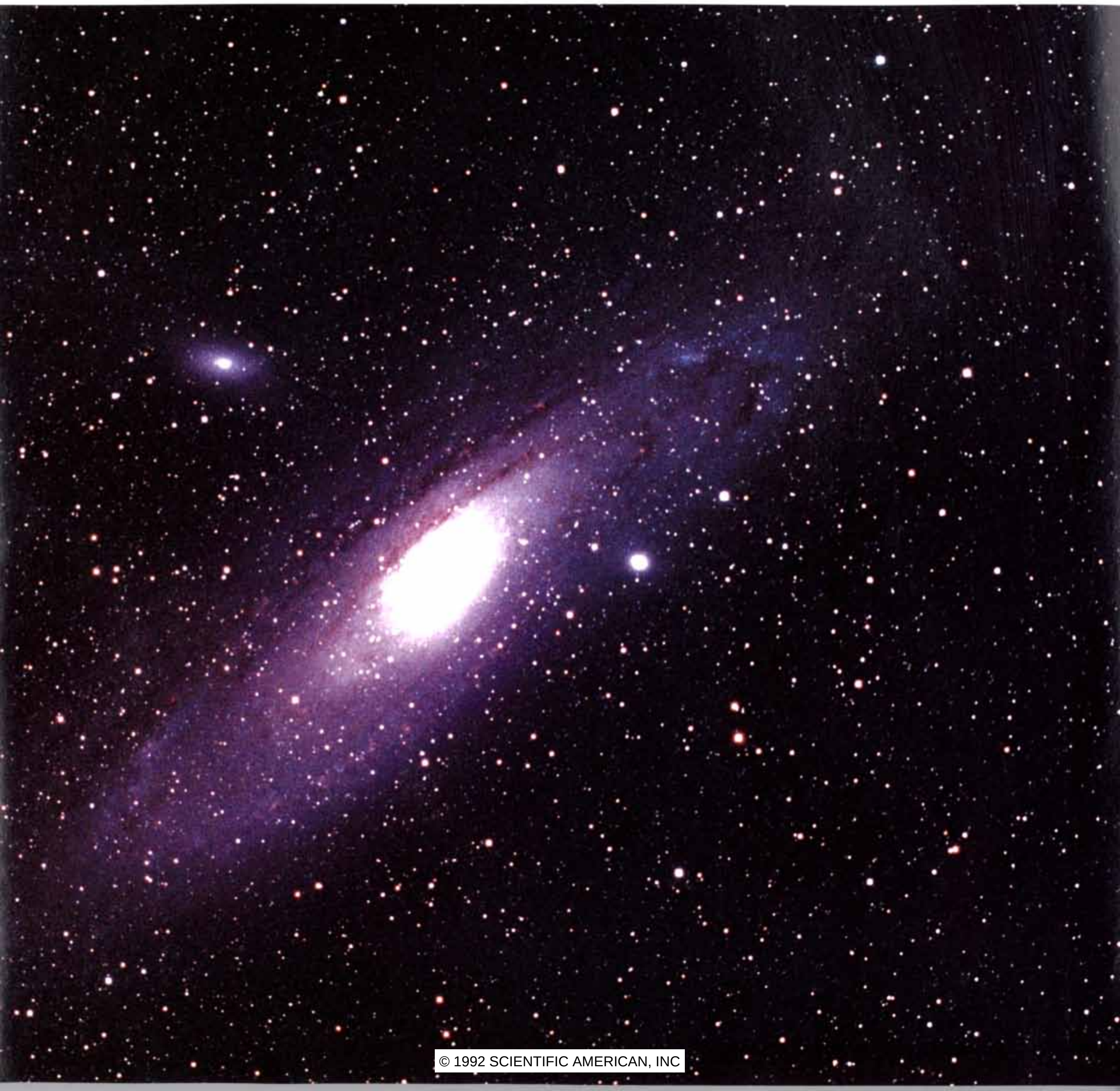
ANDROMEDA GALAXY is a prime example of why calculating the expansion rate of the universe is difficult. Andromeda is 2.5 million light-years away from the earth, but it still feels the gravitational pull of our own galaxy. Consequently, its relative motion has little to do with the expansion of the universe. By observing more distant galaxies, astronomers can detect the expansion, but they do not know its precise rate because it is difficult to measure distances to remote galaxies.

axies are moving away from the earth at tremendous velocities. Moreover, he made the remarkable observation that the velocity of recession is proportional to the distance to the galaxy. His observations provided the first evidence that the entire universe is expanding.

Hubble was the first to determine the expansion rate. Later this quantity became known as the Hubble constant: the recession velocity of the galaxy divided by its distance. A very rough estimate of the Hubble constant is 100 kilometers per second per megaparsec. (Astronomers commonly represent distances in terms of megaparsecs, where one megaparsec is the distance light travels in 3.26 million years.) Thus, a typical galaxy at a distance of 50 megaparsecs moves away at about 5,000 kilometers per second. A galaxy at 500 megaparsecs therefore moves at about 50,000 kilometers per second, or more than 100 million miles per hour!

For six decades, astronomers have hotly debated the precise value of the expansion rate. Hubble originally obtained a value of 500 kilometers per second per megaparsec (km/s/Mpc). After Hubble's death in 1953, his protégé Allan R. Sandage, also at Carnegie, continued the program of mapping the expansion of the universe. As Sandage and others made more accurate and extensive observations, they eventually revised Hubble's original value downward considerably into the range of 50 to 100 km/s/Mpc, indicating a universe far larger than suggested by the earliest measurements.

During the past two decades, new estimates of the Hubble constant have continued to fall within this same range, but preferentially toward the two extremes. Notably, Sandage and his longtime collaborator Gustav A. Tammann of the University of Basel have argued for a value of 50 km/s/Mpc, whereas Gérard de Vaucouleurs of the University of Texas has advocated a value of 100 km/s/Mpc. The controversy has created an unsatisfactory situation in which scientists have been free to choose any value of the Hubble constant between the two extremes.



© 1992 SCIENTIFIC AMERICAN, INC

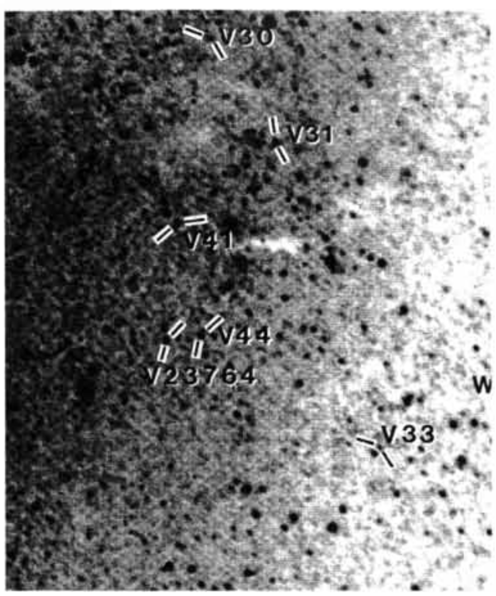
Why Cepheid Variables Pulsate

A Cepheid variable is a relatively young star, several times more massive than the sun, whose luminosity changes in a periodic way: a Cepheid brightens and then dims more slowly. It pulsates because the force of gravity acting on the atmosphere of the star is not quite balanced by the pressure of the hot gases from the interior of the star.

The imbalance occurs because of changes in the atmosphere of a Cepheid. An important ingredient in the atmosphere is singly ionized helium (that is, helium atoms that have lost a single electron). As radiation flows out of the interior of a Cepheid, singly ionized helium in the atmosphere absorbs and scatters radiation, and it may become doubly ionized (that is, each helium atom releases a second electron). Consequently, the atmosphere becomes more opaque, making it difficult for radiation to escape from the atmosphere. This interaction between radiation and matter generates a pressure that pushes out the atmosphere of the star. As a result, the Cepheid increases in size and brightness.

Yet as the atmosphere expands, it also cools, and at lower temperatures the helium returns to its singly ionized state. Hence, the atmosphere allows radiation to pass through more freely, and the pressure on the atmosphere decreases. Eventually, the atmosphere collapses back to its initial size, and the Cepheid returns to its original brightness. The cycle then repeats.

Astronomers have predicted the behavior of Cepheids with great accuracy using theoretical models of the evolution of the interior of stars as well as simulations of the flow of radiation in stars. Astronomers have confidence in Cepheids as distance indicators because they understand the underlying physics of these young stars and have observed them in great detail.



SEVERAL CEPHEID VARIABLES are apparent in the galaxy M33, a member of our own Local Group. Individual Cepheids are marked by a number and the letter V. Each dark point represents a star, whereas the white irregular patches are regions filled with dust.

In principle, determination of the Hubble constant is simple, requiring only a measurement of distance and velocity. Although measuring the velocity of a galaxy is straightforward, however, gauging the distance is rather difficult. To obtain the velocity, astronomers disperse the light coming from a galaxy and record its spectrum. The spectrum of a galaxy contains discrete spectral lines. These occur at characteristic wavelengths caused by emission or absorption by specific elements in the gas and stars that make up the galaxy. For a galaxy moving away from the earth, the positions of these spectral lines are shifted to longer wavelengths by an amount proportional to the velocity—an effect known as redshift.

To determine the distance to a galaxy, astronomers have a choice of a variety of complicated methods. Each has its advantages, but none, it seems, is perfect.

Astronomers can most accurately measure the distances to nearby galaxies by monitoring a type of star commonly referred to as a Cepheid variable. Over time, the star changes in brightness in a periodic and distinctive way. During the first part of the cycle, its luminosity increases very rapidly, whereas during the remainder of the cycle, the luminosity of the Cepheid decreases slowly. On average, Cepheid variables are about 10,000 times brighter than the sun.

Remarkably, the distance to a Cepheid can be calculated from its period (the length of its cycle) and its average apparent brightness (its luminosity as observed from the earth). In 1908 Henrietta S. Leavitt of Harvard College Observatory discovered that the period of a Cepheid is very tightly correlated with its brightness. She found that the longer the period, the brighter the star. This relation arises from the fact that the brightness of a Cepheid is proportional to its surface area. Large, bright Cepheids pulsate over a long period just as, for example, large bells resonate at a low frequency (long period).

By observing the variations in luminosity of a Cepheid over time, astronomers can obtain its period and average apparent luminosity and thereby calculate its absolute luminosity (that is, the apparent brightness the star would have if it were a standard distance of 10 parsecs away). Furthermore, they know that the apparent luminosity decreases as the distance it travels increases. Therefore, the distance to the Cepheid can be computed from the ratio of the absolute brightness to the apparent brightness.

Cepheids are useful distance indicators for many reasons. In particular, their cyclic behavior and high luminosity make them relatively easy to find and to measure.

In the 1920s Hubble used Cepheid variables to establish that other galaxies existed far beyond the Milky Way. While studying photographs of the Andromeda nebula, also known as M31, Hubble identified faint starlike images whose brightnesses varied slightly over time. He was able to show that their behavior matched that of nearby Cepheid variables. By measuring the apparent brightnesses and periods of the Cepheids in M31 he deduced that M31 was located more than several hundred thousand light-years away from the sun, well outside the Milky Way. From the 1930s to the 1960s, Hubble, Sandage and others made a tremendous effort to discover Cepheids in nearby galaxies. They succeeded in measuring the distances to about a dozen galaxies, thereby improving the foundations for deriving the Hubble constant.

One of the major difficulties with the Cepheid method is that the apparent luminosity can be diminished by the dust found between stars. The particles absorb, scatter and redden the light from all types of stars. The effects of the dust

© 1992 SCIENTIFIC AMERICAN, INC

are most severe for blue and ultraviolet light. It is therefore necessary that astronomers either observe Cepheids at infrared wavelengths where the effects are less significant or observe them at many different optical wavelengths so that they can assess the effects and correct for them.

To determine the distance to Cepheids, therefore, astronomers need telescopes and detectors that are very sensitive to light at a variety of wavelengths. Hubble, Sandage and their contemporaries used photographic plates that responded primarily to green and blue light and had an efficiency of less than a tenth of a percent. Today astronomers use solid-state charge-coupled devices (CCDs) made out of thin wafers of silicon. These devices can detect light of wavelengths from blue to red and are more than 50 percent efficient. When a photon strikes a CCD, it liberates electrons in the silicon, creating a detectable signal.

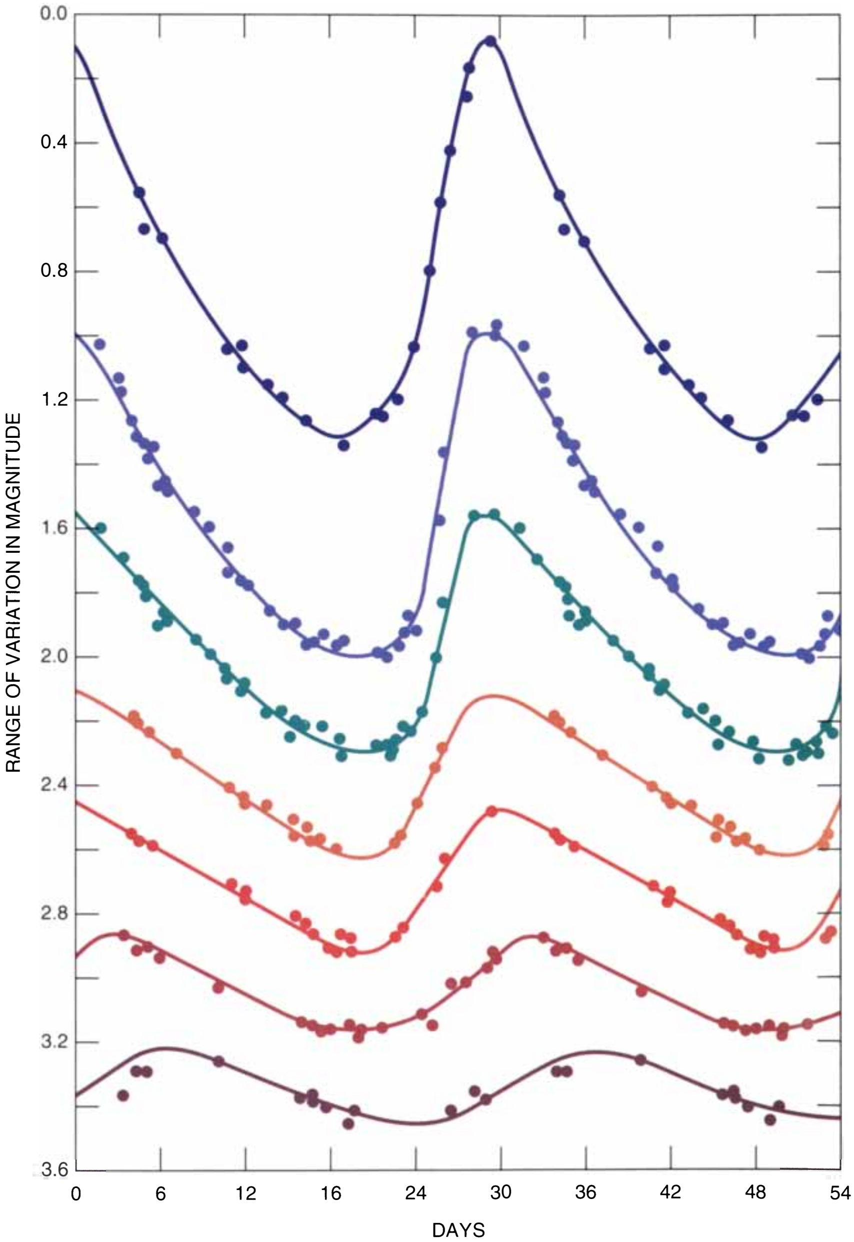
CCDs offer an enormous increase in observing efficiency over photographic plates. In addition, they record the brightness of a light source with much greater accuracy than photographic materials. The CCDs therefore make ideal detectors for studying Cepheids and for dealing with the effects of dust grains.

During the past decade, my collaborator (and husband) Barry F. Madore and I at the California Institute of Technology have carefully remeasured the distances to the nearest galaxies using CCDs and the large reflecting telescopes at many sites, including Mauna Kea in Hawaii, Las Campanas in Chile and Mount Palomar in California. As a result, we have determined the distances to nearby galaxies with much greater accuracy than ever before.

Unfortunately, the technique for measuring the distances to galaxies containing Cepheids cannot be used directly to obtain the Hubble constant. Cepheids are bright enough to be observed only in the nearest galaxies, not the distant ones. And although nearby galaxies are participating in the expansion of the universe, the gravitational interactions among the neighbors may be causing some to move much faster or slower than the rest of the universe. Consequently, to calculate the Hubble constant, astronomers must accurately determine the distances to remote galaxies, and that task is extremely difficult.

Nevertheless, astronomers have developed several methods for determining distances to remote galaxies. Because many of these techniques must be calibrated using the Cepheid distance scale, they are considered secondary distance indicators. The techniques are based either on properties of certain types of bright objects within galaxies or on characteristics of galaxies themselves. Yet scientists cannot reach a consensus about which, if any, secondary indicators are reliable. Furthermore, they disagree about how they should apply many of the methods and then whether they should adjust the results to account for various effects that might bias the results. Differences in the choice of secondary methods are at the root of almost all of the current debate about the Hubble constant.

One of the most promising techniques for measuring great distances relies on a correlation between the brightness of a galaxy and the rate at which it rotates. High-luminosity galaxies are typically more massive than low-luminosity galaxies, and so bright galaxies rotate more slowly than dim galaxies. Although the existence of such a correlation was known for some time, it was not until 1977 that R. Brent Tully of

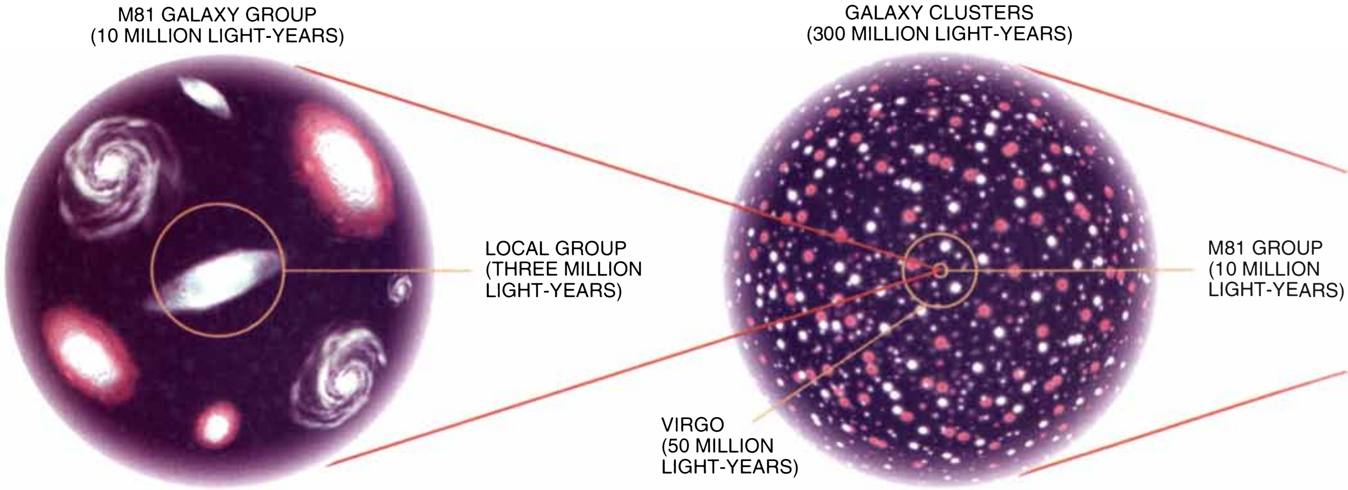


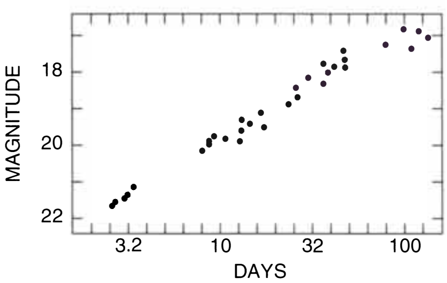
MAGNITUDE OF A CEPHEID varies cyclically over a period of days. Each colored line corresponds to observations at different wavelengths of radiation (from ultraviolet to near infrared). The amplitude of the light variation is largest toward blue and ultraviolet wavelengths. Cepheids are therefore more easily discovered using detectors sensitive to blue light.

© 1992 SCIENTIFIC AMERICAN, INC

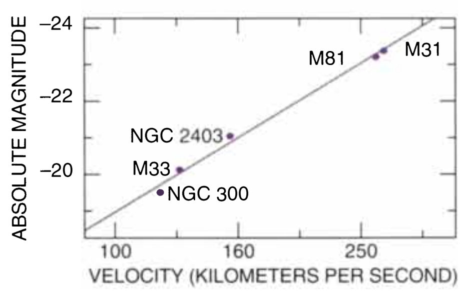
Measuring Distances to Galaxies

Astronomers can employ several different techniques for measuring distances to galaxies. Unfortunately, the accuracy of the measurements decreases as the distance to the galaxy increases. By observing stars known as Cepheid variables through ground-based telescopes, astronomers have accurately determined the distances to galaxies as far away as the M81 group, some 10 million light-years away. Using the same technique and the *Hubble Space Telescope,* they may be able to measure the distance to the Virgo cluster, approximately 50 million light-years away. By measuring the brightness of a galaxy and the velocity at which it rotates, astronomers can currently determine the distance to galaxies some 300 million light-years away. Another promising





PERIOD of a Cepheid variable—the time it takes for the star to complete one cycle of brightening and dimming—is related to its magnitude, or brightness. The measurements were made at infrared wavelengths on Cepheids in the Large Magellanic Cloud.



VELOCITY at which a galaxy rotates is related to its magnitude. This so-called Tully-Fisher relation is extremely precise, enabling accurate relative distances to be obtained. Cepheid distances to these galaxies were determined using detectors called CCDs.

the University of Hawaii and J. Richard Fisher of the National Radio Astronomy Observatory used the correlation extensively to measure distances.

The Tully-Fisher relation yields the most accurate distance measurements when observations of the brightness of a galaxy are made at infrared wavelengths. There are two reasons. First, the stars that dominate the luminosity of galaxies emit most of their radiation at near-infrared wavelengths. Second, as infrared radiation travels through space, it scatters less at longer wavelengths. Just over a decade ago the use of the Tully-Fisher relation at infrared wavelengths was pioneered by the late Marc Aaronson of the University of Arizona, Jeremy R. Mould of Caltech, John P. Huchra of Harvard University and Gregory D. Bothun of the University of Oregon. Since then, several independent groups have tested the Tully-Fisher method extensively. Most important, they have shown that the relation does not appear to depend on environment; more specifically, it is the same in the dense parts of rich clusters, in the outer parts of such clusters and for relatively isolated galaxies.

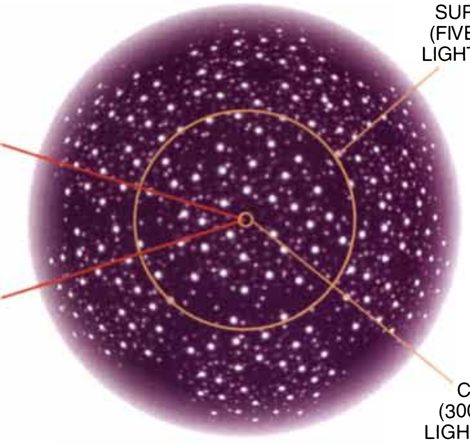
For these reasons and others, astronomers generally agree (but by no means universally accept) that the Tully-Fisher relation is one of the most accurate secondary distance indicators available. It can be used to estimate distances as far away as 300 million light-years. Another advantage of the method is that the Cepheid technique can be used to calibrate the Tully-Fisher method. A disadvantage is that astronomers currently lack a detailed theoretical understanding of the Tully-Fisher relation.

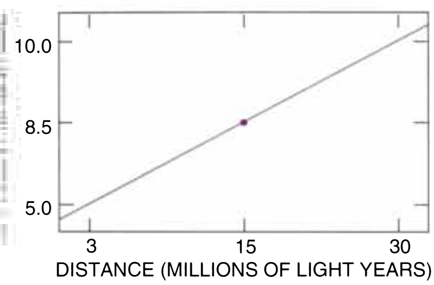
Researchers have recently developed two other distance measuring techniques [see "Mirroring the Cosmos," by Corey S. Powell; Scientific American, November 1991]. The first method, devised by George Jacoby of the National Optical Astronomy Observatories and his colleagues, involves objects known as planetary nebulae. These objects are formed when stars that are about as massive as the sun approach the end of their life cycle.

Jacoby and his co-workers found that

© 1992 SCIENTIFIC AMERICAN, INC

technique is based on the apparent peak brightness of a kind of exploding star known as a type Ia supernova. In principle, such explosions could be detected out to a distance of about half of the visible universe. The supernova technique is far less accurate than the Cepheid method, as the graphs below suggest.





PEAK ABSOLUTE BRIGHTNESS of a type Ia supernova, theory predicts, is constant, and its apparent brightness is therefore related to its distance from the earth. Yet astronomers have been able to make only one measurement to calibrate this distance scale.

the luminosities of planetary nebulae do not exceed a well-defined, upper limit. To determine the distance to a galaxy, they simply measured the apparent luminosities of the brightest planetary nebulae in that galaxy. To calibrate their method, they used galaxies with distances determined by Cepheids. They found that this technique produces distance measurements that agree very well with the Tully-Fisher method in cases where both methods have been applied.

The second method, developed by John L. Tonry of the Massachusetts Institute of Technology and his colleagues, exploits the fact that nearby galaxies appear grainy, whereas remote galaxies are more uniform in their surface brightness distribution. The graininess decreases with distance because the task of resolving individual stars becomes increasingly difficult. Hence, the distance to a galaxy can be gauged by how much the apparent brightness of the galaxy fluctuates over its surface. After determining the distances to galaxies using the surface brightness technique, Tonry compared the results with those obtained using the planetary nebula and Tully-Fisher methods and found excellent agreement. Considering the uncertainties that have plagued measurements of extragalactic distances, Tonry's findings are extremely encouraging. Yet both methods currently have only small numbers of Cepheid calibrators available.

Another distance indicator that has great potential is a particular kind of supernova known as type Ia. Supernovae are catastrophic explosions that mark the death of certain kinds of stars. Type Ia supernovae, astronomers believe, occur in double star systems in which one of the stars is a very dense object known as a white dwarf. The explosion is triggered when mass from the companion star is transferred to the white dwarf. Because supernovae release tremendous amounts of radiation, astronomers should be able to observe supernovae perhaps as far away as five billion light-years, that is, a distance spanning a radius of half the visible universe.

Type Ia supernovae make good distance indicators because at the peak of their brightness, they all are believed to produce roughly the same amount of light. Using this information, astronomers can infer their distance.

Unfortunately, supernovae are very rare events, making both their discovery and especially their calibration extremely difficult. Because they occur so infrequently, the chance of a type Ia supernovae occurring in a galaxy near enough where Cepheids can also be measured is very low. In fact, it was only during this past year that Sandage and his colleagues obtained, for the first time, a direct distance to a galaxy known to have harbored a type Ia supernova. To do so, Sandage's team made observations of Cepheids using the *Hubble Space Telescope.* Although their work represents a major advance, a single result is still insufficient to calibrate the supernova distance scale accurately.

Two other methods for determining the Hubble constant also deserve mention because they are completely independent of the Cepheid distance scale and they can be useful for measuring distances on vast cosmological scales. Moreover, preliminary applications of each of these methods currently favor a lower value of the Hubble constant.

The first of these alternative methods relies on an effect called gravitational lensing: if light from some distant source travels near a galaxy on its way to the earth, the light can be deflected. The light takes many different paths around the galaxy, some shorter, some longer, and consequently arrives at the earth at different times. If the brightness of the source varies in some distinctive way, the signal will be seen first in the light that takes the shortest path and will be observed again, some time later, in the light that traverses the longest path. The difference in the arrival times reveals the difference in length between the two light paths. By applying a theoretical model of the mass distribution of the galaxy, astronomers can calculate a value for the Hubble constant.

The second method makes use of a phenomenon known as the Sunyaev-Zel'dovich (SZ) effect. When photons from the microwave background travel through galaxy clusters, they can gain energy as they scatter off the hot plasma (x-ray) electrons found in the clusters. The net result of the scattering is a decrease in the microwave background toward the position of the cluster. By comparing the microwave and x-ray distributions, a distance to the cluster can be inferred. To determine the distance, however, astronomers must also know the average density of the electrons, their distribution and their temperature, and they must have an accurate measure of the decrement in the temperature of the microwave background. By calculating the distance to the cluster and measuring its recessional velocity, astronomers can then obtain the Hubble constant.

The SZ method and the gravitational-lensing technique are promising but have not yet been tested rigorously. To assess the uncertainties in these techniques, researchers must find more objects with the required characteristics.

The debate continues as to the best method for determining distances to remote galaxies. Consequently, astronomers hold many conflicting opinions about what the best current estimate is for the Hubble constant. Sandage and

© 1992 SCIENTIFIC AMERICAN, INC

HENRIETTA S. LEAVITT of Harvard College Observatory found, in 1908, a correlation between the period of a Cepheid variable and its absolute brightness. This correlation allows astronomers to measure distances to the nearest galaxies.

his collaborators have reported a preliminary value of 45 km/s/Mpc using the type Ia supernova method. The SZ method and the gravitational-lensing technique also support a low value for the Hubble constant.

My colleagues and I derived a best estimate by using the most recent Cepheid measurements individually to calibrate the infrared Tully-Fisher relation, the planetary nebula technique and also the surface-brightness fluctuation method. These three independent techniques give results that are in excellent agreement: they yield a high value for the Hubble constant, of about 80 km/s/Mpc.

Our measurements and those of our colleagues have many implications for the age, evolution and fate of the universe. A low value for the Hubble constant implies an old age for the universe, whereas a high value suggests a young age. In particular, a Hubble constant of 100 km/s/Mpc indicates the universe is about seven to 10 billion years old (depending on the amount of matter in the universe and the corresponding deceleration caused by that matter). A value of 50 km/s/Mpc suggests, however, an age of 15 to 20 billion years.

And what of the ultimate fate of the universe? If the average density of matter in the universe is low, as current observational estimates indicate, the current standard cosmological model predicts that the expansion of the universe will continue forever. Nevertheless, theory suggests that the universe contains more mass than that which can be attributed to luminous matter. A very active area of cosmological research is the search for this additional "dark" matter in the universe. To answer the question about the ultimate fate of the universe unambiguously, however, cosmologists require not only a knowledge of the Hubble constant and the average mass density of the universe but also an independent measure of the age of the universe. These three quantities are needed to specify uniquely the geometry and the evolution of the universe.

If the Hubble constant turns out to be high, it would have profound implications for our understanding of the evolution of galaxies and the universe. A Hubble constant of 80 km/s/Mpc yields an age estimate of from eight to 12 billion years (allowing for uncertainty in the value for the average density of the universe). These estimates are all shorter than what theoretical models suggest for the age of old stellar systems known as globular clusters. Globular clusters are believed to be among the first objects to form in our galaxy, and their age is estimated to be between 13 and 17 billion years. Obviously, the ages of the globular clusters cannot be older than the age of the universe itself.

The age estimates for globular clusters are often cited as a reason for preferring, a priori, a low value for the Hubble constant and therefore an older age for the universe. Some astronomers argue, however, that the theoretical models of globular clusters on which these estimates depend may not be complete and may be based on inaccurate assumptions. For instance, the models rely on knowing the precise ratios of certain elements present in globular clusters, particularly oxygen and iron. Moreover, accurate ages require accurate measures of luminosities of globular cluster stars, which in turn require accurate measurements of the distances to the globular clusters. Considering that both the measurements of the Hubble constant and the models and distances for globular clusters may contain errors, astronomers cannot easily judge the seriousness of the apparent age discrepancy between the universe and globular clusters.

A high value for the Hubble constant raises another potentially serious problem: it disagrees with standard theories of how galaxies are formed and distributed in space. For example, the theories make predictions about how much time is required to form the large-scale clustering that has been observed in the distribution of galaxies. If the Hubble constant is large (that is, the universe is young), the models cannot reproduce the observed distribution of galaxies.

The controversy over the value for the Hubble constant may be settled as astronomers continue to use the *Hubble Space Telescope*. In April 1990 the telescope named in honor of Edwin Hubble was launched into orbit around the earth [see "Early Results from the Hubble Space Telescope," by Eric J. Chaisson; SCIENTIFIC AMERICAN, June]. The space telescope has the potential to resolve individual stars at distances 10 times farther than can be done from the ground. It therefore provides the opportunity to discover Cepheids in a volume of space 1,000 times larger than the region that can be routinely accessed from ground-based telescopes.

One of the highest priorities of *Hubble* is to discover Cepheids in galaxies as far away as the Virgo cluster (about 50 million light-years). The Cepheids could be used to determine the distances to those galaxies, thereby calibrating various secondary distance indicators. Such observations could greatly improve measurements of the Hubble constant.

Some time ago several colleagues and I were awarded sufficient observing time on *Hubble* to find new Cepheids in more distant galaxies. In 1991 we began our first observations of the nearby galaxy M81. We identified more than 20 new Cepheids and obtained a spectacular set of light curves. Unfortunately, the telescope has not performed as expected because of the spherical aberration of the primary mirror, and most of the program has been delayed until the telescope optics have been corrected (currently planned to occur in December 1993).

Scientists have many reasons to be optimistic that the upcoming decade may allow us to resolve the current controversy over the age of the universe and to chart the course of its evolution. But the history of science suggests that we are unlikely to be the last generation to wrestle with these challenges.

FURTHER READING

MAN DISCOVERS THE GALAXIES. R. Berendzen, R. Hart and D. Seeley. Science History Publications, 1976.

THE COSMOLOGICAL DISTANCE LADDER: DISTANCE AND TIME IN THE UNIVERSE. Michael Rowan-Robinson. W. H. Freeman and Company, 1985.

THE CEPHEID DISTANCE SCALE. Barry F. Madore and Wendy L. Freedman in *Publications of the Astronomical Society of the Pacific*, Vol. 103, No. 667, pages 933–957; September 1991.

LONELY HEARTS OF THE COSMOS: THE SCIENTIFIC QUEST FOR THE SECRET OF THE UNIVERSE. Dennis Overbye. HarperCollins Publishers, 1991.

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

The Risks of Software

Programming bugs have disrupted telephone service and delayed shuttle launches. An inherent uncertainty in reliability may mean limiting a computer's role, especially in systems where software is critical for safety

by Bev Littlewood and Lorenzo Strigini

Most of us have experienced some kind of problem related to computer failure: a bill mailed in error or a day's work destroyed by some mysterious glitch in a desktop computer. Such nuisances, often caused by software faults, or "bugs," are merely inconvenient when compared with the consequences of computer failures in critical systems. Software bugs caused the series of large-scale outages of telephone service in the U.S. A software problem may have prevented the Patriot missile system from tracking the Iraqi Scud missile that killed 28 American soldiers during the Gulf War. Indeed, software faults are generally more insidious and much more difficult to handle than are physical defects.

The problems essentially arise from complexity, which increases the possibility that design faults will persist and emerge in the final product. Conventional engineering has made great strides in the understanding and control of physical problems. Although design faults are sometimes present in material products that do not contain computers, the relative simplicity of such machines has made design reliability less serious than it has become for software. The concerns expressed here, incidentally, go far beyond exotic military and aerospace products. Complex software is finding critical roles in more mundane areas, such as four-wheel steering and antilock braking in automobiles.

In this article we examine some major reasons for the uncertainty concerning software reliability and argue that our ability to measure it falls far short of the levels that are sometimes required. In critical systems, such as the safety systems of a dangerous chemical plant, it may be that the appropriate level of safety will be guaranteed only if the role of software is limited.

In theory at least, software can be made that is free of defects. Unlike materials and machinery, software does not wear out. All design defects are present from the time the software is loaded into the computer. In principle, these faults could be removed once and for all. Furthermore, mathematical proof should enable programmers to guarantee correctness.

Yet the goal of perfect software remains elusive. Despite rigorous and systematic testing, most large programs contain some residual bugs when delivered. The reason for this is the complexity of the source code. A program of only a few hundred lines may contain tens of decisions, allowing for thousands of alternative paths of execution (programs for fairly critical applications vary between tens and millions of lines of code). A program can make the wrong decision because the particular inputs that triggered the problem had not been used during the test phase, when defects could be corrected. The situation responsible for such inputs may even have been misunderstood or unanticipated: the designer either "correctly" programmed the wrong reaction or failed to take the situation into account altogether. This type of bug is the most difficult to eradicate.

BEV LITTLEWOOD and LORENZO STRIGINI collaborate as members of the project Predictably Dependable Computing Systems (PDCS), a research effort on computer dependability that brings together researchers from European countries. Littlewood received his Ph.D. in computer science and statistics from the City University in London, where he is the director of the Centre for Software Reliability. Strigini, a researcher at the Institute for Information Processing (IEI) of Italy's National Research Council (CNR), received his *laurea* in electronic engineering from the University of Pisa.
The work of the authors in this field was partially funded by the Commission of the European Communities as part of the PDCS project. The authors thank the other workers in PDCS, discussions with whom have been valuable in forming the views expressed here.

PATRIOT MISSILE streaks over Tel Aviv to intercept an incoming Iraqi Scud mis-

© 1992 SCIENTIFIC AMERICAN, INC

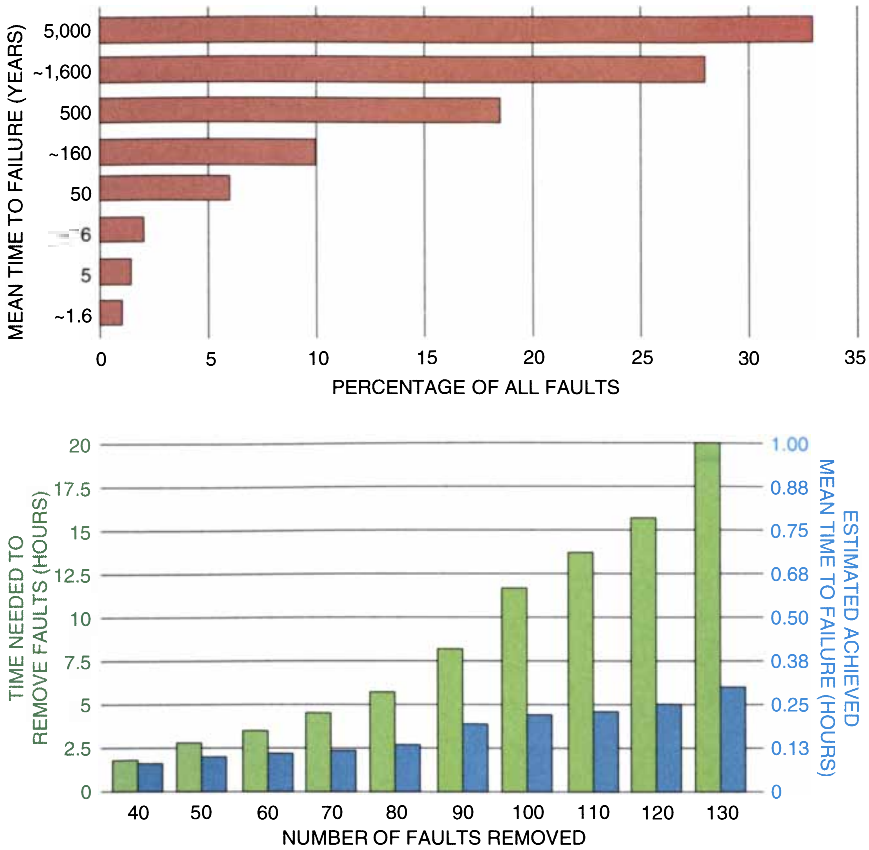
In addition, specifications often change during system development, as the intended purpose of the system is modified or becomes better understood. Such changes may have implications that ripple through all parts of a system, making the previous design inadequate. Furthermore, real use may still differ from intended purpose. Failures of Patriot missiles to intercept Scud missiles have been attributed to an accumulation of inaccuracies in the internal time-keeping of a computer. Yet the computer was performing according to specifications: the system was meant to be turned off and restarted often enough for the accumulated error never to become dangerous. Because the system was used in an unintended way, a minor inaccuracy became a serious problem.

The intrinsic behavior of digital systems also hinders the creation of completely reliable software. Many physical systems are fundamentally continuous in that they are described by "well-behaved" functions—that is, very small changes in stimuli produce very small differences in responses. In contrast, the smallest possible perturbation to the state of a digital computer (changing a bit from 0 to 1, for instance) may produce a radical response. A single incorrect character in the specification of a control program for an Atlas rocket, carrying the first U.S. interplanetary spacecraft, *Mariner 1,* ultimately caused the vehicle to veer off course. Both rocket and spacecraft had to be destroyed shortly after launch.

In all other branches of engineering, simplicity and gradual change constitute the main elements of trustworthy design. But in software engineering the unprecedented degrees of novelty and

sile during the 1991 Gulf War. On some occasions the software controlling the Patriot's tracking system may have failed, preventing the missiles from locating and destroying their Scud targets. One such failure led to the deaths of 28 U.S. soldiers.

© 1992 SCIENTIFIC AMERICAN, INC



SOFTWARE FAULTS persist even in well-debugged programs. Edward N. Adams of IBM found that bugs that remained in a system were primarily "5,000-year" bugs—that is, each of them would produce a failure only once in 5,000 years (*top*). Such faults make debugging an exercise in diminishing returns: in the test of a military command-and-control system (*bottom*), the time needed to remove the bugs begins to outpace by far the resulting improvement in the estimated reliability, measured in terms of estimated achieved mean time to failure. For visual clarity, the graphs have been plotted on different time scales.

flexibility that programming affords tempt workers to ignore these principles. Entirely new applications can be designed with apparent ease, giving a false sense of security to developers and clients who are not familiar with problems specific to software. Even the addition of novel features to a program may produce unexpected changes in existing features.

The problems of embedding complex decision rules in a design and forecasting the behavior of complex discontinuous systems are not limited to software. Designers of highly complex digital integrated circuits encounter similar problems. Software, however, is still the predominant medium for embodying extremely complex, specialized decision rules.

In addition to unintentional design bugs, flaws deliberately introduced to compromise a system can cause unacceptable system behavior. The issue of computer security, privacy and encryption requires special consideration that lies beyond the scope of this article [see "Achieving Electronic Privacy," by David Chaum; SCIENTIFIC AMERICAN, August].

Given that perfect software is a practical impossibility, how can we decide whether a program is as reliable as it is supposed to be? First, safety requirements must be chosen carefully to reflect the nature of the application. These requirements can vary dramatically from one application to another. For example, the U.S. requires that its new air-traffic control system cannot be unavailable for more than three seconds a year. In civilian airliners, the probability of certain catastrophic failures must be no worse than 10^{-9} per hour.

In setting reliability requirements for computers, we must also take into account any extra benefits that a computer may produce, because not using a particular system may itself incur harm. For example, military aircraft are by necessity much more dangerous to fly than are civilian ones. Survival in combat depends on high performance, which forbids conservative design, and a new computer system may improve the airplane's chances even if it is less safe than computers used in commercial airplanes. Similarly, in the design of a fly-by-wire civilian aircraft, such as the Airbus A320 or the Boeing 777, the possibility that software may cause accidents has to be weighed against the likelihood that it may avoid some mishaps that would otherwise be caused by pilot error or equipment failure.

We believe that there are severe restrictions on the levels of confidence that one can justifiably place in the reliability of software. To explain this point of view, we need to consider the different sources of evidence that support confidence in software. The most obvious is testing: running the program, directly observing its behavior and removing bugs whenever they show up. In this process the reliability of the software will grow, and the data collected can now generally be used, via sophisticated statistical extrapolation techniques, to obtain accurate measures of how reliable the program has become.

Unfortunately, this approach works only when the reliability requirements are fairly modest (say, in the range of one failure every few years) when compared with the requirements often set for critical applications. To have confidence at a level such as 10^{-9} failure per hour, we would need to execute the program for very many multiples of 10^9 hours, or 100,000 years. Clearly, this task is not possible. In the time spans for which it is feasible to test, assurance of the safety would fall many orders of magnitude short of what is needed.

The problem here is a law of diminishing returns. When we continue debugging a program for a very long time, eventually the bugs found are so "small" that fixing them has virtually no effect on the overall reliability or safety. Edward N. Adams of the IBM Thomas J. Watson Research Center empirically analyzed "bug sizes" over a worldwide data base that involved the equivalent use of thousands of years of a particular software system.

The most extraordinary discovery was that about a third of all bugs found were "5,000-year" bugs: each of them produced a failure only about once in 5,000 years of execution (the rates from other bugs varied by several orders of magnitude). These rare bugs made up a sizable portion of all faults because bugs that caused higher failure rates were encountered, and so removed, earlier. Eventually, only the 5,000-year bugs will make the system unreliable, and remov-

© 1992 SCIENTIFIC AMERICAN, INC

al of one of these will bring negligible improvement in reliability.

Extrapolating from testing and debugging also implies an unsubstantiated assumption—namely, that a bug, once encountered, is simply corrected. In reality, an attempt to fix a bug sometimes fails. It may even introduce an entirely novel fault. Because nothing would be known about the new bug, its effect on the reliability of the system would be unbounded. In particular, the system might not even be as reliable as it was before the bug was found.

Therefore, a prudent course would be to discount completely the history prior to the last failure. This precaution, critically important in situations that involve safety, would require an evaluator to treat the software after the last fix as if it were a completely new program. Only the most recent period of error-free working would influence judgment about the safety of the program. But even this conservative course of action cannot provide much confidence. Our research has shown that under quite plausible mathematical assumptions, there is only about a 50–50 chance that the program will function without failure for the same length of time as it had before.

The problem of estimating safety is actually even more serious. To have any confidence in the numerical results, we must subject the program to situations it might encounter in reality. This approach ensures that inputs causing failures are encountered with the same frequency with which they would in fact arise. In addition, the tester should always be able to decide whether the program's output is actually correct. The problems here are similar to those of designing and implementing the software itself. To construct an accurate test environment, we need to be sure that we have thought of all circumstances that the software will meet. Just as the unexpected often defeats us in system design, so it is in test design. It would be wise to retain an element of skepticism about the representativeness of the testing and thus about the accuracy of the figures.

The problem in demonstrating extreme reliability or safety for any individual piece of software is simply lack of the necessary knowledge. For complex software, the unpalatable truth seems to be that there are severe limitations to the confidence one can place in a program. Merely observing a program's behavior is not the way to be sure that it will function properly for 100,000 years. How else might we acquire such confidence?

An obvious prerequisite for high reliability is that software be built with methods that are likely to achieve reliability. One method uses "formal" techniques, which rely on mathematical proofs to guarantee that a program will function according to specification. Indeed, formal techniques have become a topic of wide interest. Such methods, though currently limited by practical problems in their scope of application, can effectively avoid programming errors arising in the translation from the specification to the actual program.

Unfortunately, specifications must also be formal statements. In other words, the user's needs would have to be expressed in a mathematical language. That task is not simple: it requires a careful choice of those aspects of the real world to be described in the formal language and an understanding of both the detailed practical problems of the application and of the formal language. Errors would likely be introduced during this process, and we could not reasonably claim the program would never fail.

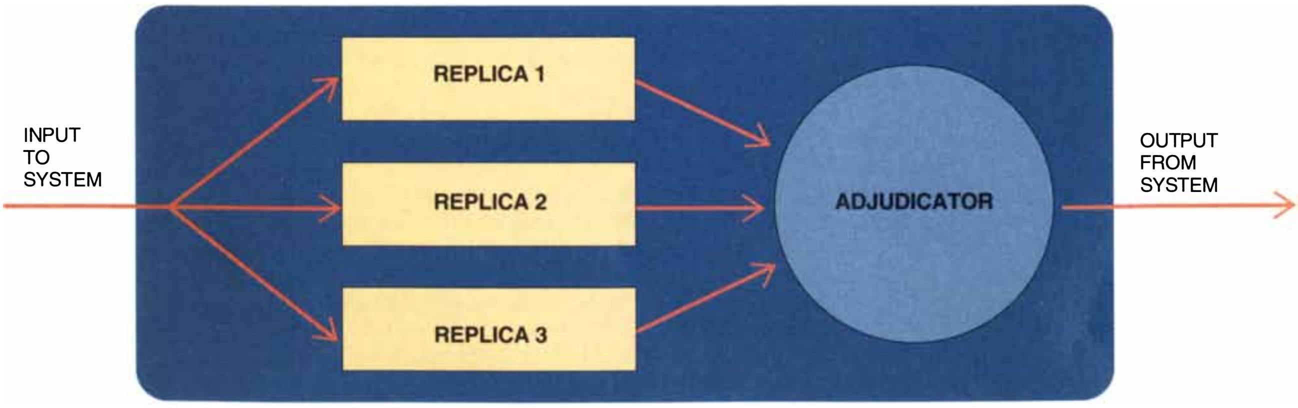
Another method now widely used (in avionic and railroad control applications, for instance) to achieve high reliability is fault tolerance, or protective redundancy. A typical way of applying redundancy is to have different design teams develop several versions of the program. The hope is that if the teams make mistakes, the errors will be different. Each version of the program provides its "opinion" of the correct output. The outputs pass to an adjudication phase, which produces a single output that would be correct if the majority of versions gave the correct result.

Some evidence exists that such design diversity delivers high reliability in a cost-effective manner. Different design teams, however, may make the same mistakes (perhaps because of commonalities in cultural background) or conceptually different mistakes that happen to make the versions fail on the same fault. The adjudicator would therefore produce incorrect output.

To measure the reliability of fault-tolerant software, it is necessary to gauge the statistical correlation between failures of the different versions. Unfortunately, the task turns out to be as hard as trying to measure the reliability by treating the whole system as a single entity—and we have seen the difficulty of doing that.

So if formal proofs do not enable one to claim that a program will never fail and if fault tolerance cannot guarantee reliability, there seems no choice but to evaluate reliability directly, using methods that are acknowledged to be of limited adequacy. How do the regulatory authorities and software users deal with this uncertainty?

There are three approaches. The first



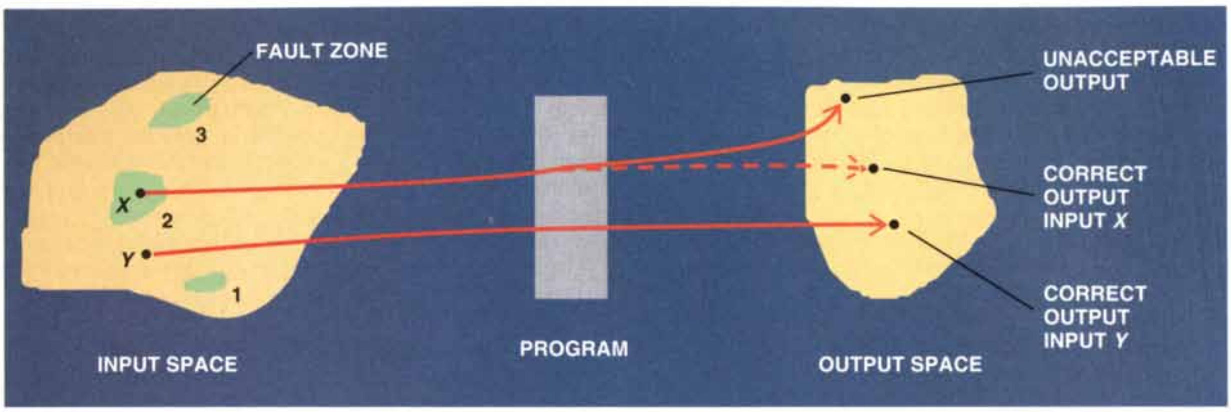
DESIGN DIVERSITY helps to increase the reliability of software systems. Each program version, or replica, is developed independently by different design teams. The "adjudicator" decides the actual output of the system by using, for example, the median value produced by the replicas or the value "voted for" by a majority. The adjudicator could be another redundant system or could consist of noncomputer technology, such as hydraulic actuators.

© 1992 SCIENTIFIC AMERICAN, INC

classifies design-caused failures as "nonquantifiable" errors and avoids specifying requirements for the software. This method is now in fairly wide use. For instance, in civil aviation, the U.S. Federal Aviation Administration Advisory Circular 25.1309-1A describes "acceptable" means for showing compliance with some federal aviation regulations. It states that catastrophic failure conditions (the worst category) must be "so unlikely that they are not anticipated to occur during the entire operational life of all airplanes of one type." The suggested quantitative expression is the probability of failure of not more than 10^{-9} per hour of flight. Software, however, is explicitly excluded from this circular, "because it is not feasible to assess the number or kinds of software errors, if any, that may remain after the completion of system design, development, and test."

The widely used document of the Radio Technical Commission for Aeronautics, RTCA/DO-178A, similarly avoids software measures. The document, which gives guidelines for manufacturers who must seek certification by aviation authorities, explicitly refuses to mandate quantitative terms or methods for evaluating software reliability or safety. Instead the commission regards a correct engineering approach—tight management, thorough reviews and tests, and analysis of previous errors—as more critical than quantitative methods. The basic message of RTCA/DO-178A "is that designers must take a disciplined approach to software: requirements definition, design, development, testing, configuration management and documentation." That is, the best assurance of reliability is to verify that utmost care was used in the design.

How good is such assurance? Arguably, not very good: there is no evidence that superior design and produc-



The Nature of Software Failure

Software occasionally fails because it contains design faults. Some have argued that such failures are systematic—that is, because writing software is a purely logical exercise, there is nothing intrinsically uncertain about it. If enough is known about the inputs, the program's behavior would be completely deterministic. We believe, however, that software failures cannot be mathematically described only in deterministic terms. In fact, we think that describing the nature of software failures requires a probabilistic treatment, just as we use statistics to describe how often, on average, electrical or mechanical devices fail.

To see why, consider all the possible inputs (called the input space) that the software might encounter in its life (*above*). An input for an operation of the software is a set of digital data (numbers) read from the outside world and from information already stored in the computer's memory. In the figure above, the input space is shown in the two dimensions of the printed page, but in practice the space would usually consist of many dimensions.

Here the input space contains three fault zones numbered 1 to 3. Input *x* lies in fault 2, that is, it would cause the program to produce an unacceptable output. On the other hand, the program can successfully execute input *y*, which does not lie in any fault zone.

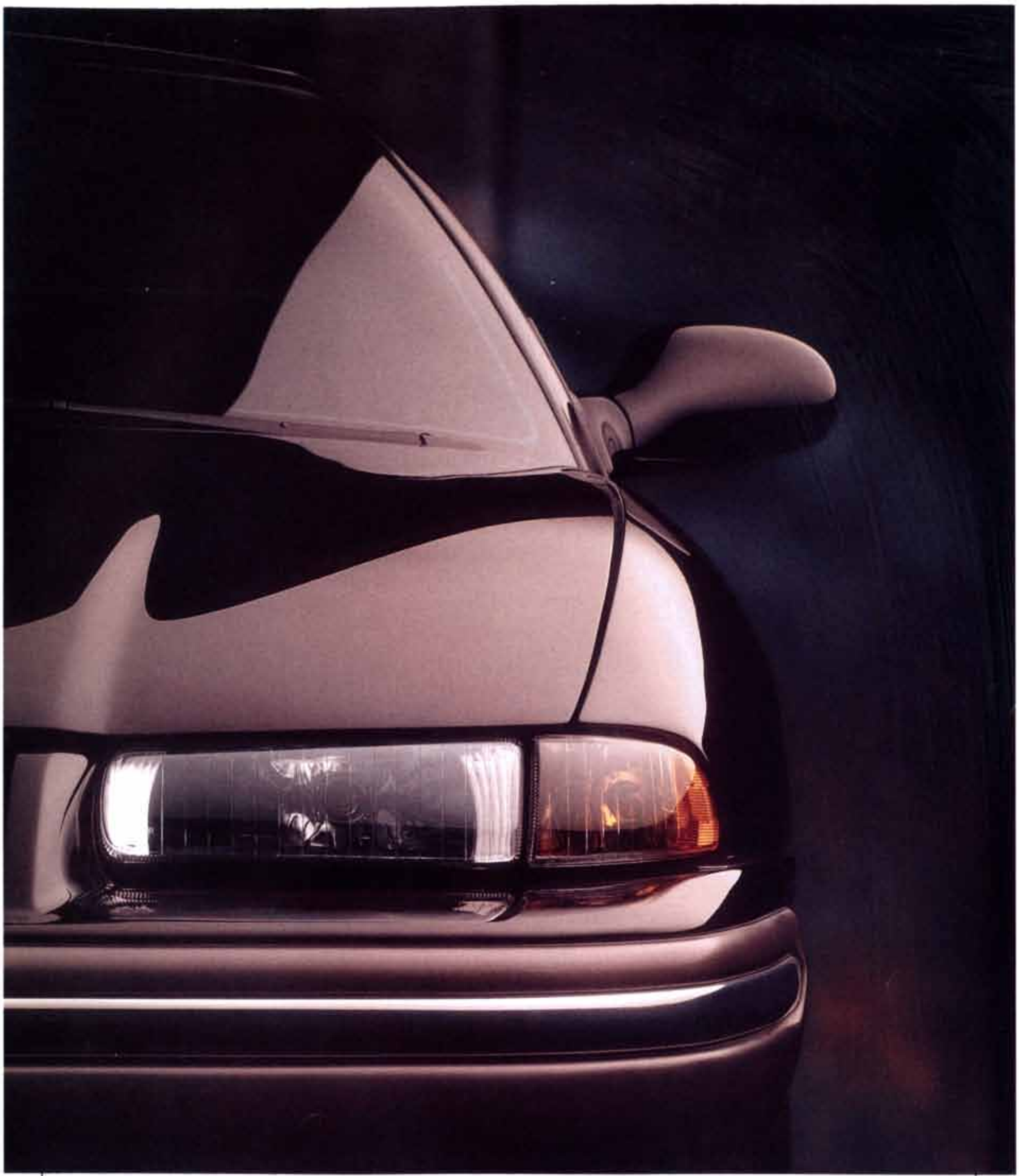
A program is tested by executing it with many inputs and checking whether the results are correct. If a right answer is produced during testing, it will also be produced whenever the same input is presented. For most programs, testing for all possible inputs would require billions of billions of years—hence, the need to infer failure probabilities from testing on a sample of inputs.

We would like to know when the program will next fail, but that is not possible because of the inherent uncertainty in the process. First, uncertainty arises from the physical mechanisms that determine the succession of inputs (called the trajectory in the input space). We can never be sure which inputs will be selected in the future, and different inputs will have different chances of being selected. Second, we are uncertain about the sizes and locations of the fault regions in the input space. Even if we knew the trajectory, we would still not know when the program would encounter a fault.

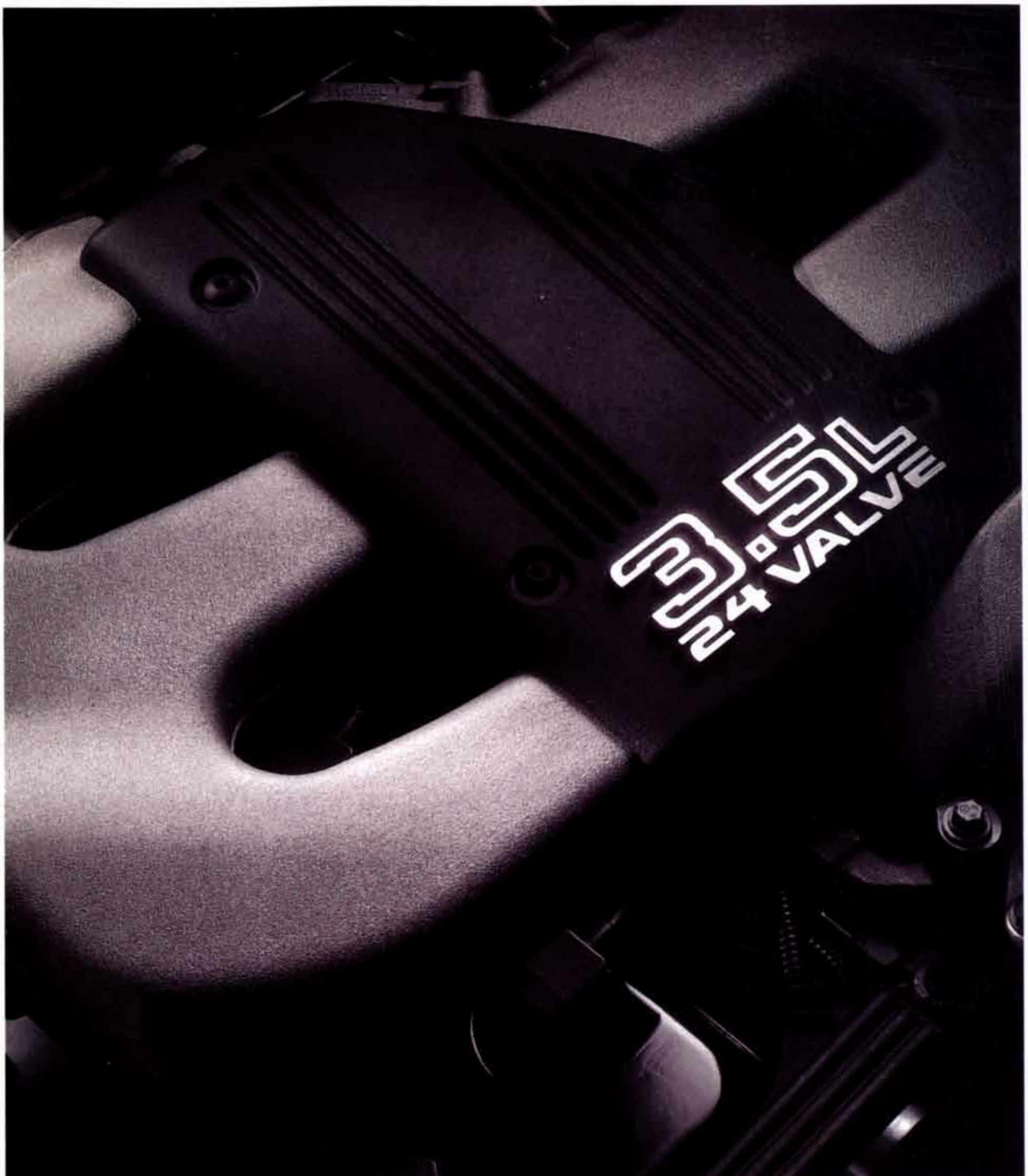
Therefore, we must describe our belief about the future failure behavior of the program in terms of probabilities. We might ask what is the probability that we can survive a particular number of inputs before failure. Or we might ask what the probability is that a randomly selected input causes a failure. Both questions can often easily be turned into a time-based measure of reliability—that is, the probability that the program will execute perfectly for a particular length of time.

In conclusion, we are forced to consider the process of successive failures of a program to be just as "random" as that of a hardware device. The use of a probability-based reliability measure is therefore inevitable.

© 1992 SCIENTIFIC AMERICAN, INC



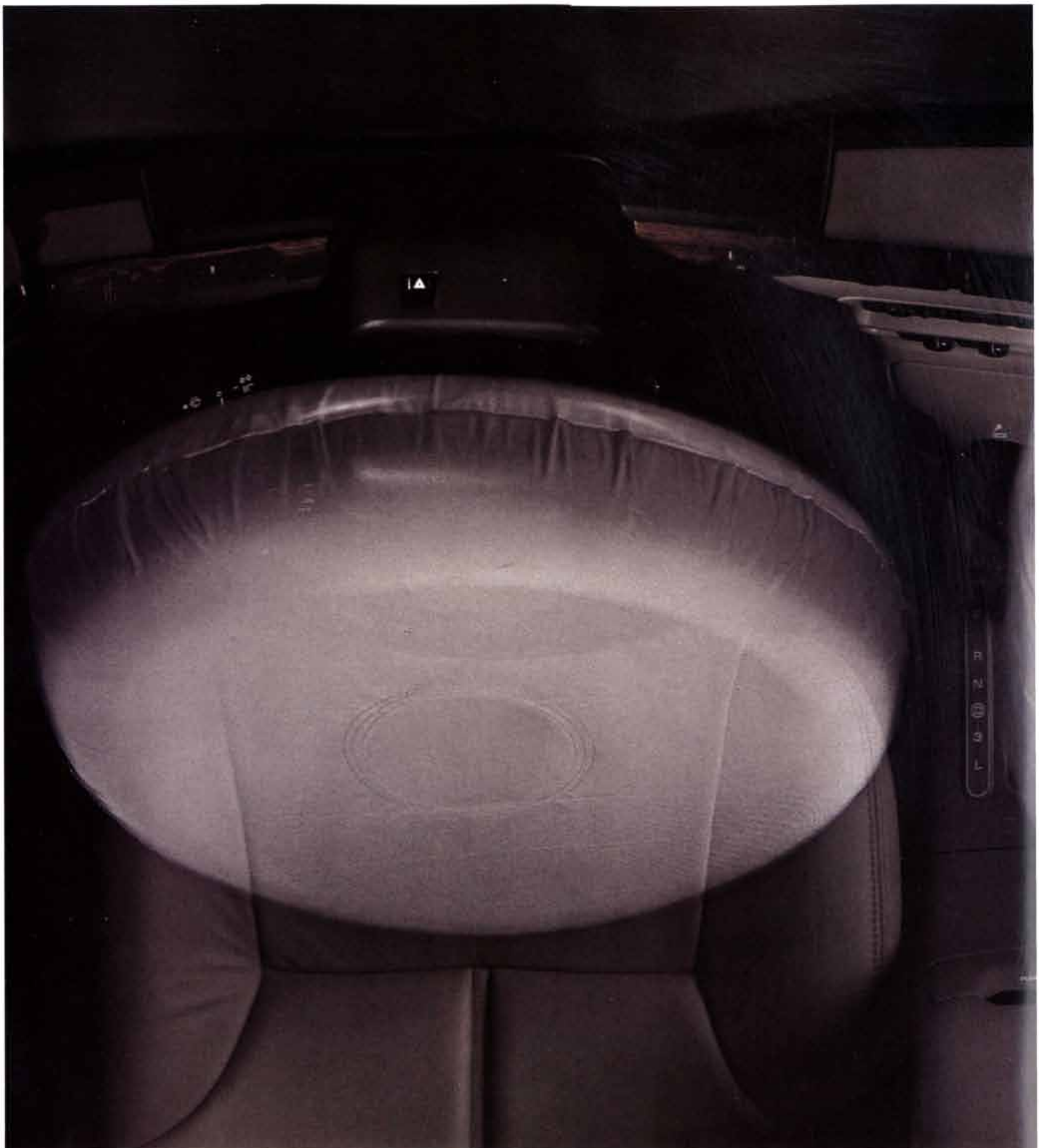
© 1992 SCIENTIFIC AMERICAN, INC



It has more horsepower than the Acura Legend, and BMW 525i, and more torque than Nissan 300ZX. It's German, right?

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC



© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

Where this car is made is not nea
Introducing the C

As *AutoWeek* magazine puts it, this car "has traits that render meaningless such adjectives as European, Japanese and American." It's world class. Which simply means it can compete with anything the world has to offer, regardless of national origin. Its 3.5-liter, 24-valve overhead cam V-6 makes it the rival of any four-door sedan in any showroom in the country. When it comes to handling, *AutoWeek* adds: "Chrysler didn't just target what was out there. It anticipated where the world would be and aimed beyond that mark. It hit where it aimed." In the realm of safety, the Chrysler Concorde has driver *and* passenger air bags and anti-lock four-wheel disc

See limited warranties at dealer. Restrictions apply. 3/36 excludes normal maintenance, adjustments and wear items. *MSRP example. Title, taxes and destination fee extra

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

The Chrysler Concorde.
Nothing less than the renaissance
of the American car.†

© 1992 SCIENTIFIC AMERICAN, INC

tion methods consistently yield superior products. We cannot even be certain whether the best current methods ever produce sufficient reliability for the more demanding applications.

Eschewing the quantification of software safety poses a serious limitation for many potentially dangerous systems, especially those that require an overall probabilistic risk assessment before operation. Probabilities can be predicted with reasonable accuracy for physical failures caused by stress and wear. But this accuracy cannot be used in efforts to assess the risk that the entire system (that is, hardware and software) will fail if nothing more precise can be said other than that the best effort was made to avoid mistakes. Simply mandating the use of "best practice" does not solve the problem. We hasten to add that it would be foolish to abandon techniques known to improve reliability and safety just because we do not know exactly how much they help. Standards that encourage their use are certainly beneficial, but they do not solve the problem of knowing that the software has the required safety.

The second—and we think, better—approach would require that the system be designed so the role of software in it is not too critical. "Not too critical" here means that the required software reliability is sufficiently modest so the reliability can be demonstrated before the system is deployed. This approach has been taken for the new Sizewell B nuclear reactor in the U.K., where only a 10^{-4} probability of failure on demand is needed from the software-based protection system.

There are well-established methods for limiting the criticality of any one component. For example, an industrial plant whose operations are controlled primarily by computers may be equipped with safety systems that do not depend on any software or other complex design. A safety or backup system usually performs simpler functions than does the main control system, so it can be built more reliably. Safety is possible if the backup systems are completely separated from the main systems. They could be built with different technology or use alternative sensors, actuators and power sources. Then the probability that both primary and backup (or safety) systems will fail simultaneously may be justifiably considered low.

The third approach is simply to accept the current limitations of software and live with a more modest overall system safety. After all, society sometimes demands extremely high safety for what may be irrational reasons. Medical systems are a good example. Surgeons are known to have fairly high failure rates, and it would seem natural to accept a computerized alternative if the device is shown to be as good as or only slightly better than the human physician. Indeed, in the near future robotic surgeons will probably perform operations that are beyond the capabilities of humans.

SIZEWELL B will be the first nuclear reactor in the U.K. to contain both conventional and software-based protection systems for emergency shutdowns. Critics argue that the complexity of the software system—it relies on hundreds of microprocessors and more than 100,000 lines of code—makes it difficult to ensure reactor safety.

The three approaches to regulating software safety may seem rather disappointing. Each sets limits on either the degree of safety in the system or the amount of complexity in the program. Perhaps the only way to learn more about the necessary compromises between safety and complexity is to study the failures (or lack thereof) of software in operation.

Unfortunately, there is a paucity of data from which to fashion statistical predictions. Information on software failure is seldom made public. Companies fear that sharing such knowledge would harm their competitive stance. They worry even more that publishing it would antagonize public opinion. People might see the detection of a software fault as an indication of low production standards, even though it may actually attest to a very thorough procedure applied to very high quality software. But secrecy can only allow expectations of safety to climb to increasingly unrealistic levels. Some investigators have suggested that the government make mandatory the logging and disclosing of failure data in critical software systems. Such regulations would remove the fear that companies volunteering the information would be hurt.

However it is obtained, an extensive collection of data would in time help to quantify the efficacy of different production and validation techniques. The information would help establish more realistic rules for gauging the trustworthiness of software systems. Thus, for software that is not fully tested statistically, the acceptable claims of safety could be tied to explicit upper bounds that would depend on the complexity of the program. Such an approach might allow us to justify claims for the reliability and safety of software beyond what is now believable.

In the meantime, we should remain wary of any dramatic claims of reliability. Considering the levels of complexity that software has made possible, we believe being skeptical is the safest course of action.

FURTHER READING

EVALUATION OF SAFETY-CRITICAL SOFTWARE. David L. Parnas, A. John van Schouwen and Shu Po Kwan in *Communications of the ACM*, Vol. 33, No. 6, pages 636–648; June 1990.

SOFTWARE SAFETY IN EMBEDDED COMPUTER SYSTEMS. Nancy G. Leveson in *Communications of the ACM*, Vol. 34, No. 2, pages 34–46; February 1991.

FORUM ON RISKS TO THE PUBLIC IN COMPUTERS AND RELATED SYSTEMS. Moderated by Peter Neumann. Available as the usenet newsgroup comp.risks, or by request on the internet from risks-request@csl.sri.com.

RISKS TO THE PUBLIC IN COMPUTERS AND RELATED SYSTEMS. Regular column edited by Peter Neumann in *Communications of the ACM*.

© 1992 SCIENTIFIC AMERICAN, INC

Visualizing Biological Molecules

Computer-generated images are aiding research in molecular structure and helping to elucidate the complex chemistry of life

by Arthur J. Olson and David S. Goodsell

"The eye, which is called the window of the soul, is the chief means whereby the understanding may most fully and abundantly appreciate the infinite works of nature." The words of Leonardo da Vinci eloquently capture the intimate relation between vision and comprehension. Yet modern science often confronts objects that are invisible to the human eye. Chemists and biochemists in particular have been thwarted by the fact that they cannot see the molecules they endeavor to study. The atomic details of molecules cannot be discerned even through electron microscopes.

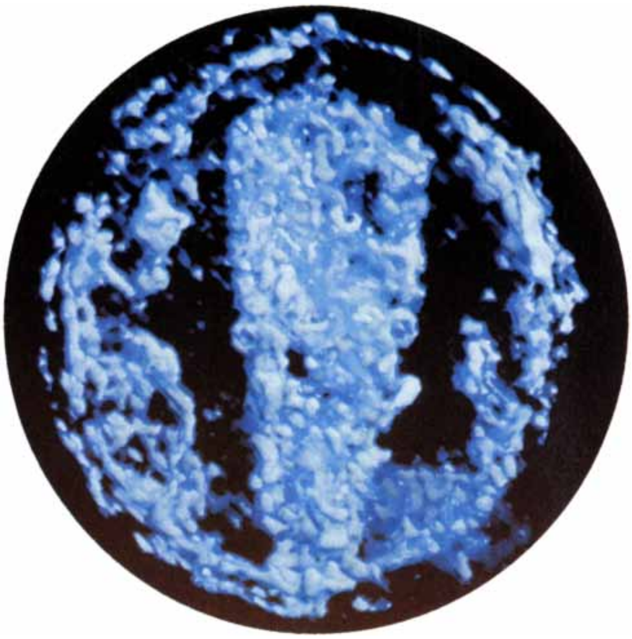
In recent years, however, computer technology has made it possible to simulate convincing, scientifically accurate pictures of molecules. Such images allow biochemists and molecular biologists to explore, in a familiar visual way, the complex molecules built by cells. Computer graphics help to disclose, for example, how antibodies seek out foreign molecules and how enzymes provide exactly the right environment to initiate a chemical reaction. A clear picture of the structure of a molecule can carry great conceptual weight. One such image—the diagram of the double-helix shape of DNA published by James Watson and Francis Crick—revolutionized understanding of human heredity and genetic disease.

Scientists gather the raw data for molecular images in several ways. X-ray crystallography is currently the most successful. A researcher irradiates a crystal composed of a particular molecule with an intense beam of x-rays, which scatter into a distinctive pattern. The pattern is mathematically analyzed to reveal the spatial distribution of electrons and, by extension, the location of every atom in the molecule.

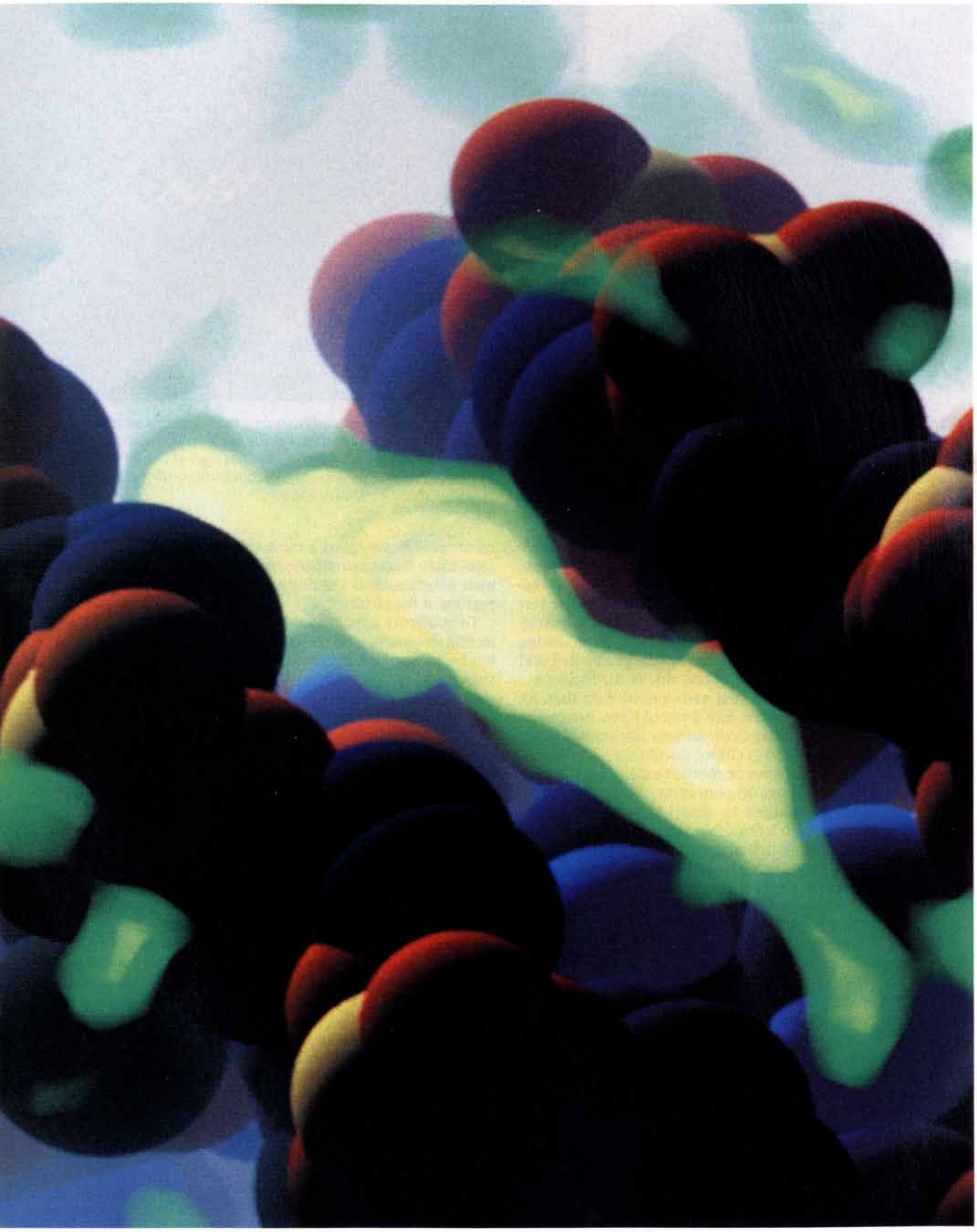
Nuclear magnetic resonance (NMR) spectroscopy offers an alternative approach to determining the structure of a molecule. A solution containing the molecule of interest is placed in a powerful magnetic field. The sample is then exposed to pulses of radio waves; nuclei of certain atoms in the molecule respond by emitting their own radio waves at frequencies determined by their local chemical environments. These frequencies are interpreted to disclose the approximate distances between atoms in the molecule. By combining those constraints with the known chemical properties of the molecule, one can deduce the positions of the constituent atoms.

ARTHUR J. OLSON and DAVID S. GOODSELL are working to expand the role of computer graphics for studying the function and structure of large biological molecules. Olson received his Ph.D. from the University of California, Berkeley, in 1975 and conducted postdoctoral research at Harvard University, where he used x-ray crystallography to study the structure of viruses. In 1981 he founded the Molecular Graphics Laboratory at the Research Institute of Scripps Clinic in La Jolla, Calif. Olson's biochemical films and molecular images have appeared in many popular and technical settings. Goodsell received his Ph.D. in 1987 from the University of California, Los Angeles, where he also studied x-ray crystallography. He then joined Olson's laboratory, developing molecular rendering techniques and computational methods for drug design. Goodsell recently returned to U.C.L.A. to continue his work in crystallography.

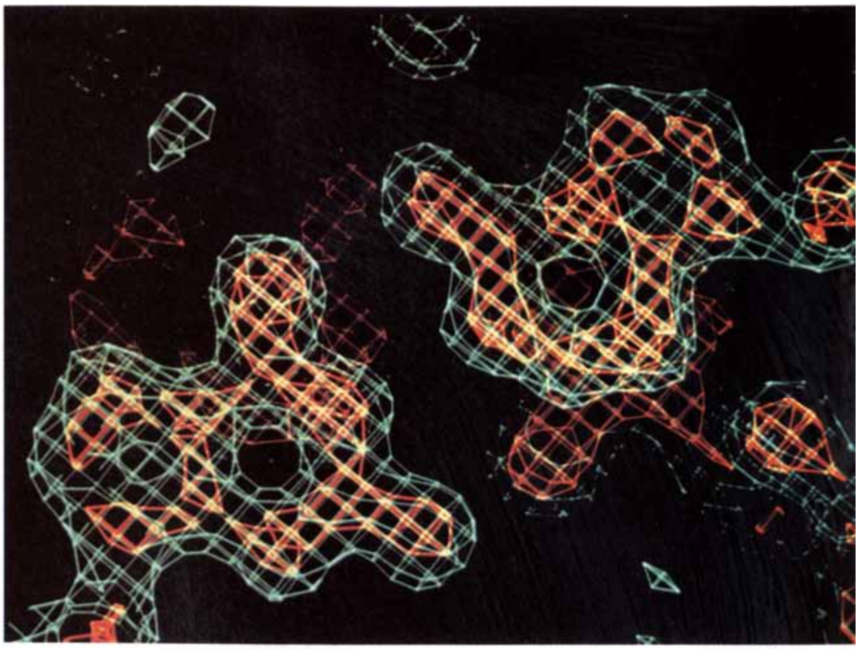
SIMULATED IMAGES of the molecular world were created by means of computer graphics. A picture of the human immunodeficiency virus (*below*), based on electron microscope data from U. Skoglund of the Karolinska Institute and S. Höglund of Uppsala University, shows a cone-shaped core containing genetic material surrounded by a spherical envelope. A view of a drug binding to DNA (*right*) was drawn using x-ray crystallographic data collected by R. E. Dickerson of U.C.L.A. The drug appears as a region of high electron density (*green and yellow*) filling the narrow groove of the DNA (*dark spheres*).



© 1992 SCIENTIFIC AMERICAN, INC



© 1992 SCIENTIFIC AMERICAN, INC



X-RAY CRYSTALLOGRAPHY provides the information needed to simulate a picture of a molecule. A crystal of DNA scatters x-rays in a characteristic pattern (*left*). Computer analysis of the pattern can reveal the distribution of electrons in each DNA molecule. A cage of lines drawn around regions of high electron density (*right*) illustrates the locations of the atoms in the molecule; an unusual guanine-adenine mispair is seen here. A pixel-based image (*opposite page*) gives a better feeling for the three-dimensional arrangement of the atoms but requires more time to compute.

Within the past few years, materials scientists have developed a third method for observing the atoms in molecules, called scanning probe microscopy. A molecule is immobilized on a flat surface, and a needle whose tip is only a few atoms wide is scanned across the surface. A feedback loop allows the tip to follow the exact contour of each atom, tracing out its shape. Repeated passes of the needle gradually build up a three-dimensional contour of one side of the molecule.

All three techniques yield vast amounts of data that are far easier to interpret if recast into a visual form. Before the widespread use of computers, researchers laboriously sifted through information on strip charts, oscilloscopes and photographs and then built brass or plastic models based on the results. Because of the huge amount of work involved, scientists were effectively limited to studying small molecules containing no more than a few dozen atoms.

Researchers interested in biological molecules, which contain hundreds to hundreds of thousands of atoms, were therefore avid users of the first computers. In 1947 Raymond Pipinsky of Pennsylvania State University developed an analog machine, XRAC, to transform his x-ray crystallographic data into an intelligible molecular picture. As computers have advanced, so too has the magnitude of the scientific problems to which they are applied. Modern digital computers, used in conjunction with computer graphics, enable scientists to produce detailed pictures of enormous molecules, including enzymes, antibodies and even entire viruses.

Two conceptually distinct graphic approaches are commonly used to create pictures of molecules. One builds up an image out of sets of lines drawn from point to point. The other method generates an image from a dense map of dots, or pixels. Each technique has its own advantages and drawbacks. Because a line can be described by merely two positions, line-based displays can draw and redraw an image rapidly, letting an investigator manipulate the image interactively. Images drawn on pixel-based displays (usually a color monitor) take longer to generate, because each picture element must be assigned a color value; a typical monitor contains more than one million pixels. But pixel displays can simulate effects such as shading and shadowing, which add to the realism of the pictures.

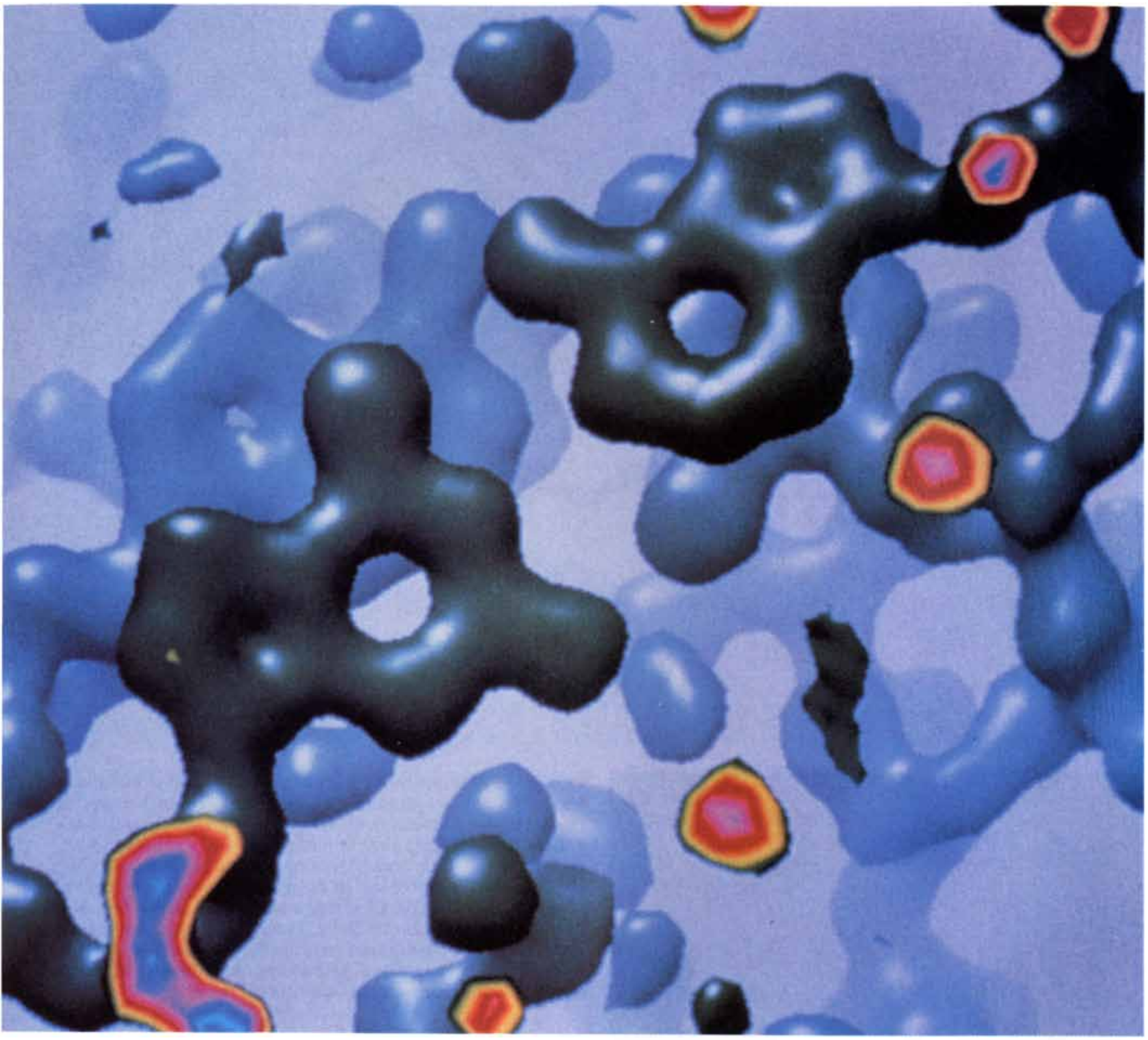
To construct a molecular image, researchers begin by collecting information on the structure of the molecule, most often by means of x-ray crystallography. X-rays scatter most strongly where the electron density is highest—that is, around the atoms in the molecule. Hence, regions that exhibit high electron densities are atoms; regions having low densities are empty space. (Electron microscopy can furnish similar but coarser three-dimensional maps of electron density that do not resolve individual atoms.)

Just as cartographers draw lines of constant elevation on a map to segregate hills and valleys, crystallographers use computer graphics to draw a boundary surface through the data, separating atoms from empty space. The surface may be portrayed as a thick mesh of lines that resembles a birdcage. Using a graphics program, the scientist then fits a chain of atoms inside the surface, following the convoluted contours indicated by the electron density data.

Pixel-based images provide a clearer view of the crystallographic results. For example, one can assign specific colors and optical properties to various values of the data. In the DNA electron density map shown on the previous page, parts of the molecule having high values of electron density are rendered opaque and colored, whereas regions of low density appear transparent. Through a process known as volume rendering, the graphics software forms an image that simulates how light would travel through an object possessing those optical properties. Unfortunately, volume-rendered images require far more time to calculate than do the line-based images. Clarity is gained at the cost of speed of manipulating the view of the molecule.

Once the coordinates of the constituent atoms are known, the computer offers a host of techniques by which to analyze a molecule. Molecular graphics can focus and simplify the pic-

© 1992 SCIENTIFIC AMERICAN, INC



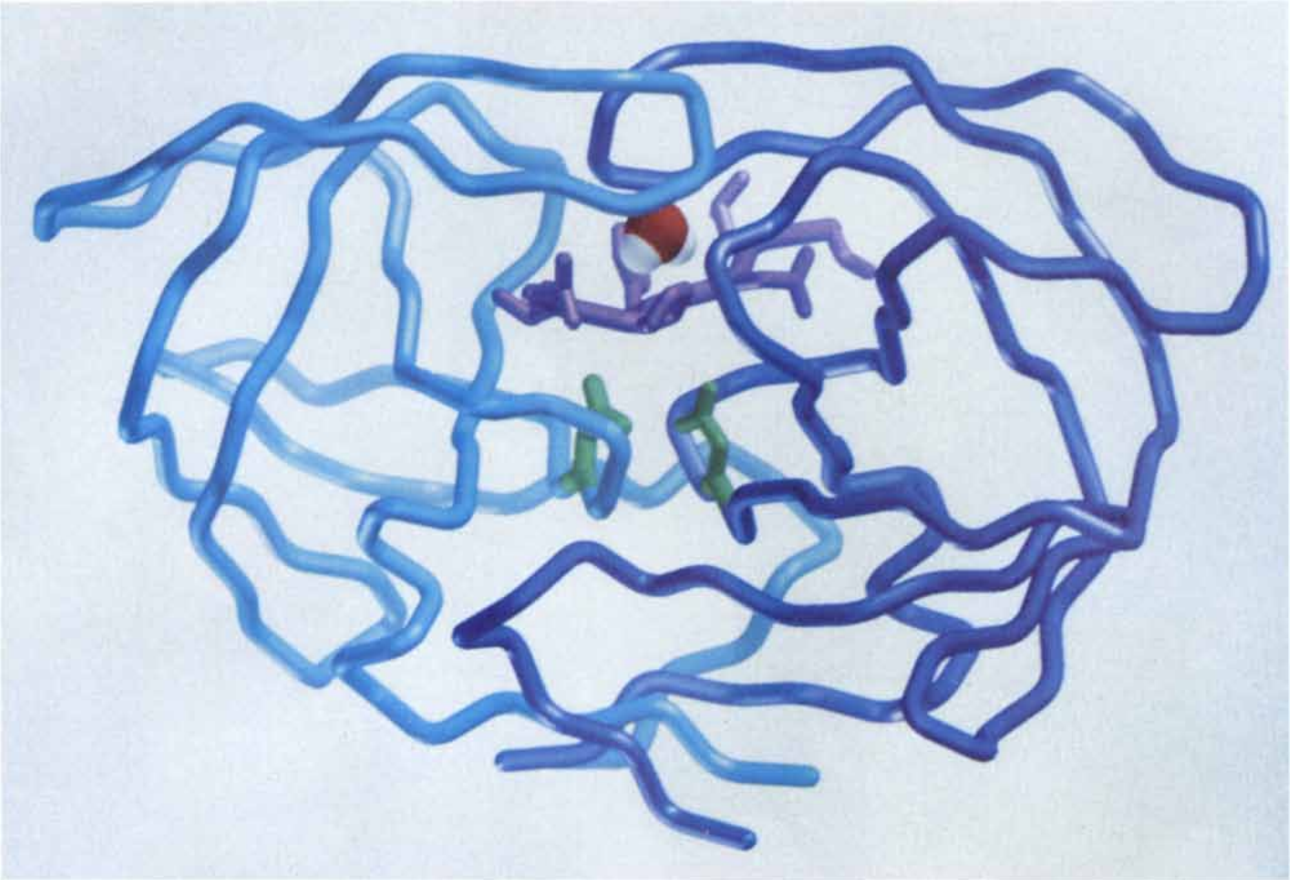
ture of the molecule while maintaining all the relevant information. Biochemists often find it enlightening to look at the way a protein chain folds into a compact molecule. Jane S. Richardson of Duke University popularized a simple but effective graphic representation that follows the overall folding of the protein but eliminates the confusing tangle of individual atoms. The resulting ribbon diagram facilitates classifying the many diverse protein structures into a limited number of distinct folding motifs.

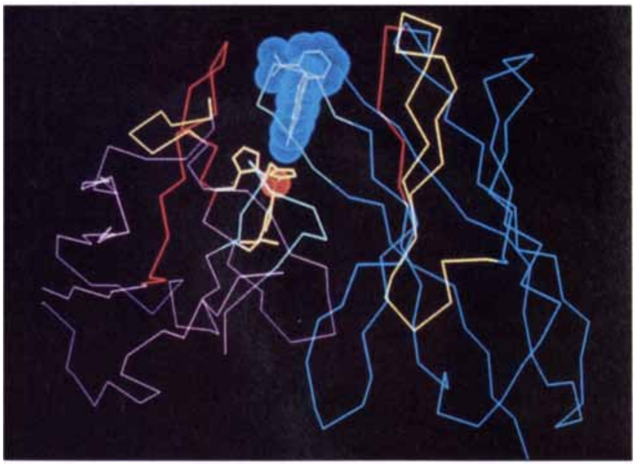
Computers also assist the study of the shape of a biological molecule, which in turn determines how it interacts with other molecules. The easiest way to depict the outer topography of a molecule is to create a space-filling diagram. In this procedure the computer draws atoms in the molecule as spheres whose radii reflect how close they may approach one another. Placing all the spheres in their proper locations renders a highly lucid image of the entire molecule. Coloring each atom according to its chemical nature conveys still more information.

Space-filling pictures show a molecule as it might appear if it were magnified to visible size. B. K. Lee, now at the National Institutes of Health, and Frederic M. Richards of Yale University took a different approach and calculated what a molecule might look like to a water molecule. They used a computer program to roll an imaginary water molecule around all sides of a molecule, noting where the water touched. In this way, they derived a picture that shows the molecule's surface but omits those regions that are sequestered from surrounding water molecules. Such a picture helps to elucidate, for instance, how proteins interact with the water always present in living systems.

The geometric arrangement of atoms is only one aspect of a molecule's nature; the chemical and physical properties of each atom—its charge, size and interactions with other atoms—are also important. Peter J. Goodford of the University of Oxford has developed a method to determine how a biological molecule chemically interacts with atoms in other molecules. He sequentially places a computer-simulated probe atom at various locations around the molecule. At each point, the computer calculates the chemical interaction between probe and molecule, yielding a catalogue of places

© 1992 SCIENTIFIC AMERICAN, INC





MOLECULAR DESIGN is a rapidly growing application of molecular computer graphics. HIV protease (*above*), a molecule crucial to the maturation of the virus, offers an attractive target for drug design. The protein backbone is shown in blue; the site of catalytic activity is green. An inhibitor drug (*purple*) binds at the active site of the protease, preventing normal viral function. The data for this model are from A. Wlodawer of the National Cancer Institute. In another experiment, workers fabricated a customized enzyme (*left*) by grafting a site to bind a metal ion (*small red sphere*) onto an antibody that binds to the chemical fluorescein (*blue dotted region*).

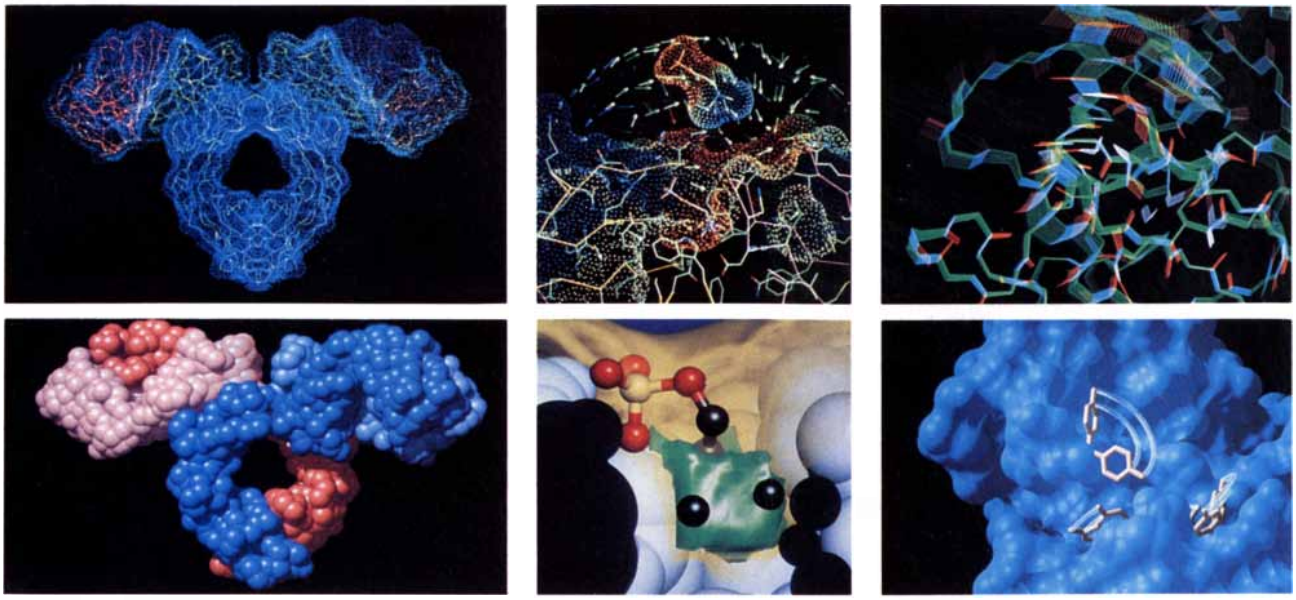
favorable and unfavorable to the probe atom. Pictures produced using this type of analysis can highlight the chemical "hot spots" where an atom is likely to bind.

Computer graphics can also capture the elusive dynamic behavior of biological molecules. In trillionths of a second, molecules vibrate, twist and rotate. Their motions are invisible to x-ray crystallography and NMR spectroscopy, because such experiments take hours or days to perform. Dynamic computer simulations, however, can follow the motions of a molecule through thousands of time steps, generating a corresponding number of snapshots of the changing structure. The researcher may then scroll through the entire simulation, stopping at leisure to examine the most interesting time steps. The best frames may be combined into a short movie that depicts the molecule's dynamics.

One of the most exciting applications of the new graphics techniques lies in computer-aided molecular design. As scientists have improved their understanding of biological molecules, they have increasingly sought to modify specific molecules to suit a particular need. Designing antibiotic drugs, constructing novel proteins and engineering useful microorganisms are just a few of the goals bioengineers are pursuing to improve human health and the quality of life.

Computer graphics are assisting in the design of drugs to treat diverse disorders, including hypertension, emphysema, glaucoma and various forms of cancer. Researchers use computer graphics to test a wide range of candidate drugs before beginning the time-consuming process of synthesis and laboratory testing. One particularly promising effort involves the design of antiviral agents to control the effects of the human immunodeficiency virus (HIV). Medical researchers have isolated a number of proteins from the virus and determined their structures. Among these proteins are HIV reverse transcriptase, the molecule that translates viral RNA into DNA so that it can incorporate itself in the host cell's DNA, and HIV protease, a crucial molecule that allows the virus to mature and to cause further infection.

© 1992 SCIENTIFIC AMERICAN, INC



MOLECULAR STRUCTURE AND FUNCTION can be studied easily, clearly and accurately using computer graphics. The top images were drawn on line-based devices, the bottom ones on pixel-based displays. Researchers may examine the shape of an antibody using a dotted surface (*top left*) or a series of solid-looking shaded spheres (*bottom left*). Computers can portray the binding of an antigen by displaying arrows that indicate the direction of the molecule's electrostatic field (*top center*) or by using volume-rendering techniques to display the strength of carbon interactions; green indicates regions favorable to carbon (*bottom center*). The extremely rapid internal motions of an antibody may be depicted as a multiple image (*top right*) or as one in a series of snapshots of physically interesting moments (*bottom right*).

Workers have succeeded in crystallizing HIV protease alone and bound with various inhibitor molecules, making it possible to study them by x-ray crystallography. Computer-based analysis of the structures of the molecules has helped identify a growing list of candidate drug compounds. Several of them appear effective in laboratory chemical tests and can arrest the growth of HIV in a cell culture. Although issues of toxicity and efficacy in actual patients remain to be solved, at least one computer-designed HIV protease inhibitor—Ro31-8959, fabricated by Hoffmann-La Roche in the U.K.—has shown sufficient promise that it is now being tested in clinical trials.

Encouraged by the many advances in the understanding of protein structure and function, Richard A. Lerner and his colleagues at the Research Institute of Scripps Clinic have embarked on a particularly ambitious project: designing customized enzymes to catalyze, or facilitate, certain chemical reactions. The researchers are modifying antibodies to act as catalysts. Antibodies possess a remarkable ability to recognize and distinguish between various molecules, so a catalytic antibody could be constructed to aid a carefully selected reaction [see "Catalytic Antibodies," by Richard A. Lerner and Alfonso Tramontano; SCIENTIFIC AMERICAN, March 1988]. Specifically designed catalytic antibodies could one day attack a virus or break up a blood clot without harming the patient's own healthy cells.

In collaboration with Victoria A. Roberts, John A. Tainer and Elizabeth D. Getzoff, also at Scripps, Lerner has modified an antibody to create a chemical site where metal atoms can bind. Computer graphics helped to guide the researchers in constructing the site. The ability to add a metal to antibodies is an important step toward the goal of tailor-made catalysts, because many reactions depend on metal atoms for catalysis.

The promise of molecular computer graphics has barely begun to be realized. The speed and memory capacity of computer hardware are doubling every 18 months, leading to commensurate improvements in the versatility of software. Virtual-reality simulators that can immerse the researcher in a tangible molecular world are themselves becoming a reality. Video-display goggles change the view in response to head motions; force-feedback mechanisms let the researcher "feel" the forces acting on the molecule in view.

In a prototype being developed at the University of North Carolina under the direction of Frederick P. Brooks, Jr., scientists can use a computerized simulator to test candidate drugs by feeling how well they fit into a target molecule. An innovative project at that same facility has linked a scanning tunneling microscope with a virtual-reality system. The goal is to enable the scientist to see and feel the atomic details of a molecule being probed by the microscope. Such systems may someday enable humans to interact with the submicroscopic world as easily as they do with the world of direct senses.

Perhaps the greatest virtue of molecular computer graphics lies in its potential to improve scientific communication. High-speed data networks will enable workers in different parts of the world to examine simultaneously the latest results in molecular research. Interactive video will permit students at all levels to study molecular structure and function. And sophisticated simulations coupled with realistic graphics will allow laypeople to obtain, for the first time, a personal feeling for the complex chemical world within themselves.

FURTHER READING

MOLECULAR MODELING SOFTWARE AND METHODS FOR MEDICINAL CHEMISTRY. N. Claude Cohen, Jeffrey M. Blaney, Christine Humblet, Peter Gund and David C. Barry in *Journal of Medicinal Chemistry*, Vol. 33, No. 3, pages 883–894; March 1990.

COMPUTER GRAPHICS: PRINCIPLES AND PRACTICE. James D. Foley, Andries van Dam, Steven K. Feiner and John F. Hughes. Addison-Wesley Publishing, 1990.

MACROMOLECULAR GRAPHICS. A. Olson and D. Goodsell in *Current Opinion in Structural Biology*, Vol. 2, pages 193–201; April 1992.

JOURNAL OF MOLECULAR GRAPHICS. Edited by W. G. Richards. Butterworth-Heinemann Publishers, quarterly.

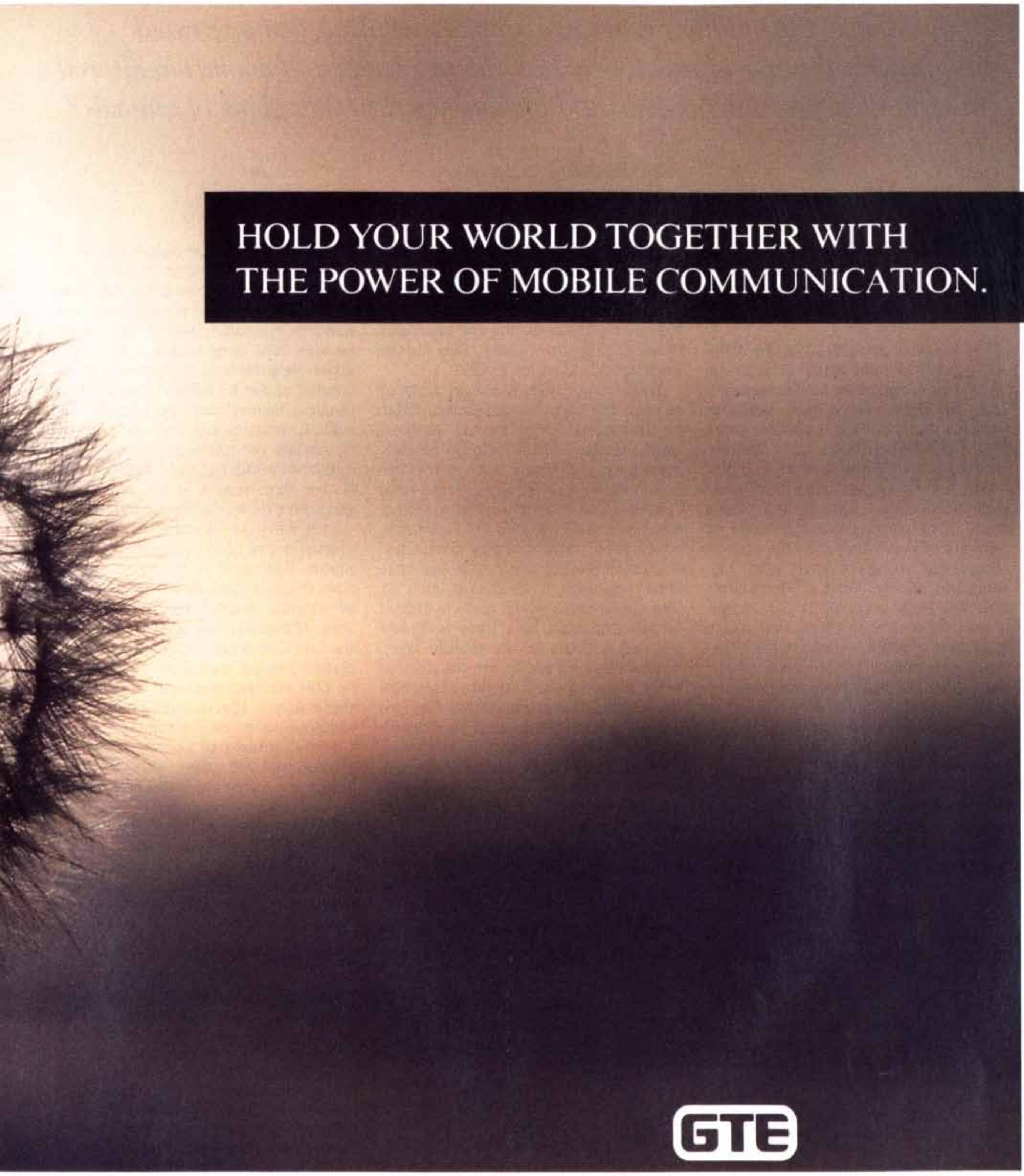
© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

metropolitan areas. From Tampa to San Francisco to the state of Hawaii. GTE's Airfone® In-flight Telephone Service is now on 15 major airlines. And Amtrak is expanding the availability of Railfone®.

Offer a comforting word to your child, check in with your office, open the door of a new business opportunity. At GTE, we give you the power to touch your world.

Because at GTE, the power is on.



GTE

THE POWER IS ON

© 1992 SCIENTIFIC AMERICAN, INC

The Big Bang of Animal Evolution

Almost 600 million years ago animal evolution demonstrated an unmatched burst of creativity. Has the mechanism of evolution altered in ways that prevent fundamental changes in the body plans of animals?

by Jeffrey S. Levinton

What has been is what will be, and what has been done is what will be done; and there is nothing new under the sun.
—Ecclesiastes 1:9

Biologists are united in the belief that the vast array of animals, plants and other life-forms populating the globe evolved from simple organisms that came into existence more than three billion years ago. The oldest fossils are of simple algae and other single-celled organisms; more complex multicellular animals and plants made their appearance hundreds of millions of years later. The increase in complexity seems to have been anything but steady. Most of evolution's dramatic leaps occurred rather abruptly and soon after multicellular organisms first evolved, nearly 600 million years ago during a period called the Cambrian. The body plans that evolved in the Cambrian by and large served as the blueprints for those seen today. Few new major body plans have appeared since that time. Just as all automobiles are fundamentally modeled after the first four-wheel vehicles, all the evolutionary changes since the Cambrian period have been mere variations on those basic themes.

JEFFREY S. LEVINTON is professor and chair of the department of ecology and evolution at the State University of New York at Stony Brook. He received a bachelor's degree from the City College of New York and a Ph.D. from Yale University, both in geology. He is a marine ecologist and evolutionary biologist, a Guggenheim Fellow and the author of the books *Marine Ecology* and *Genetics, Paleontology and Macroevolution.* Levinton also recently participated in the revision and reissue of Rachel Carson's classic *The Sea around Us.*

Evolutionary biology's deepest paradox concerns this strange discontinuity. Why haven't new animal body plans continued to crawl out of the evolutionary cauldron during the past hundreds of millions of years? Why are the ancient body plans so stable?

These major body plans are familiar to even the casual amateur naturalist. In the animal kingdom the simplest multicellular creatures are the radially symmetric cnidarians, such as jellyfish and anemones, which have bodies that consist of two layers of tissue. Somewhat more complicated are flatworms, which have three primary tissue layers, are bilaterally symmetric and have sense organs concentrated at one end. The coelomates, which include almost all other animals, have three body layers and a cavity in the middle layer. Within this vast group are the distinct body plans of the annelids (segmented worms), the echinoderms (the sea stars, sea cucumbers, starfish and other pentamerally symmetric creatures), the arthropods (insects, true bugs, spiders and crustaceans), the mollusks, the vertebrates and many less well known organisms.

Such structural differences are the basis for the hierarchical system with which biologists first began to classify animals and plants. Echinoderms, arthropods, annelids and the other groups each make up one phylum, or major division, of the kingdom Animalia; a phylum is defined by the distinctive body plan of its members. Each phylum is in turn divided successively into classes, orders and smaller groups, down to the level of species.

In 1859 Charles Darwin explained why this taxonomic hierarchy exists in nature. Evolution, he realized, is a branching process, and each division in the hierarchy represents another branch point. Phyla are distinguished by characters that reflect the oldest and deepest levels of evolutionary association.

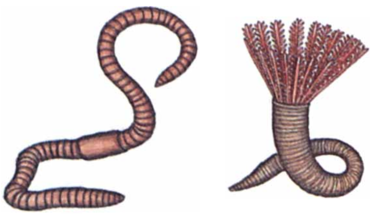
All the known animal phyla that readily fossilize appeared during the 60-million-year Cambrian period. We cannot be sure how early within it the phyla arose. Nevertheless, compared with the context of the 3.5 billion years of all biological history and the roughly 570 million years since the start of the Cambrian, the phyla do seem to have appeared suddenly and simultaneously. For that reason, some paleontologists refer to the Cambrian "explosion."

Even when we consider the taxonomic level below phyla—classes—it is apparent that most of the basic innovation occurred early. Richard K. Bambach of Virginia Polytechnic Institute and State University has shown that after the late Cambrian the number of new classes arising decreased precipitously. This evidence seems to confirm that there was a spectacular evolutionary radiation in the early Cambrian.

Some features of the Cambrian explosion are still rather uncertain. The assumption that the fossil record tells the truth about when the phyla originated is a matter of great controversy. The progenitors of the distinct animal groups found in the Cambrian could conceivably have diverged hundreds of millions of years earlier, yet they might not have left fossils because they lacked

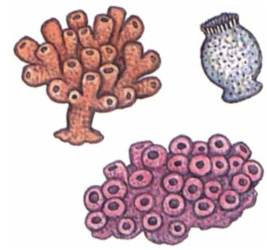
CAMBRIAN EXPLOSION was characterized by the sudden and roughly simultaneous appearance of many diverse animal forms almost 600 million years ago. No other period in the history of animal life can match this remarkable burst of evolutionary creativity. Most of the Cambrian creatures shown here were reconstructed from fossils by Simon Conway Morris and Harry Whittington of the University of Cambridge.

© 1992 SCIENTIFIC AMERICAN, INC

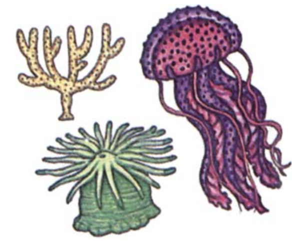


ANNELIDA
SEGMENTED WORMS

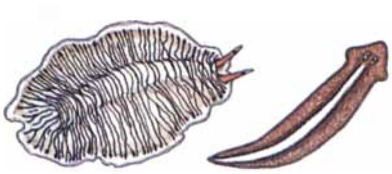
NEMATODA
ROUNDWORMS



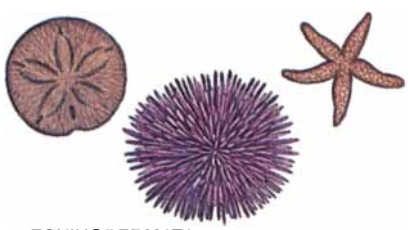
PORIFERA
SPONGES



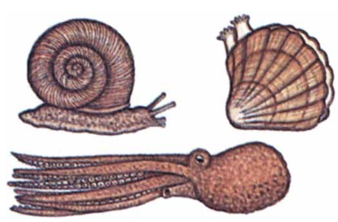
CNIDARIA
JELLYFISH, CORALS, SEA ANEMONES



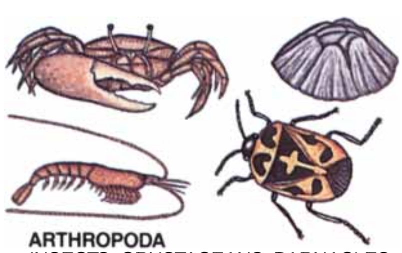
PLATYHELMINTHES
FLATWORMS



ECHINODERMATA
STARFISH, SEA URCHINS, SAND DOLLARS



MOLLUSCA
CLAMS, SNAILS, OCTOPI, SQUID



ARTHROPODA
INSECTS, CRUSTACEANS, BARNACLES

CHORDATA
FISH, AMPHIBIANS, REPTILES, BIRDS, MAMMALS

UNIQUE BODY PLANS are the hallmark of the animal phyla, the largest taxonomic groupings. All the organisms within a phylum share certain evolutionary innovations; the further division of the animals into smaller taxonomic categories, such as

shells or skeletons. If so, the Cambrian diversification might not have been as explosive as is generally assumed.

Investigators have found conflicting evidence on this point. The only known animallike fossils that predate the Cambrian belong to a peculiar group discovered in 1947 at the Ediacara Hills in southern Australia by R. C. Sprigg, a government geologist; they were first described by Martin F. Glaessner of the University of Adelaide. (Similar fossils have since been found elsewhere.) The Ediacaran fauna seem to be an evolutionary dead end, however: they cannot easily be related to living organisms or even Cambrian fossils.

Attempts to find an answer with the tools of molecular biology have been inconclusive. Biologists postulate that the sequences of nucleotide bases in DNA and of amino acids in proteins mutate at approximately constant rates; the sequences can therefore be used as a kind of molecular clock. After comparing the globin proteins in living organisms, Bruce Runnegar of the University of California at Los Angeles estimated that multicellular animals probably divided into lineages that anticipated the major phyla more than 900 million years ago—well before the Cambrian. On the other hand, evidence obtained by sequencing the 18S ribosomal RNA (molecules that aid in the synthesis of proteins) from various species suggests that many of the phyla appeared almost simultaneously, perhaps during the latter part of the pre-Cambrian era. The times of origin for the phyla and their exact relations remain obscure.

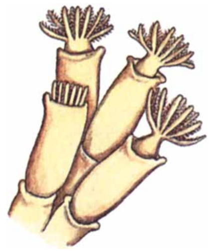
All in all, the facts still point to an explosion of complex life near the beginning of the Cambrian. The actual extent of that explosion can be appreciated only by looking critically at the fossil record. The most spectacular assemblage of Cambrian fossils comes from the Burgess Shale in British Columbia. Its remarkably well preserved specimens were first discovered in 1809 by Charles D. Walcott of the Smithsonian Institution. Although Walcott thought the strange fossils could be allied with living groups, many paleontologists now think the Burgess Shale and other Cambrian sediments contain many unique body plans that flourished early in the Cambrian, only to become extinct later. Stephen Jay Gould of Harvard University has popularized this view in his book *Wonderful Life.*

A good example of such a strange Burgess Shale fossil is *Wiwaxia,* a spiny creature about one inch long, described in great detail by the paleontologist Simon Conway Morris of the University of Cambridge. Morris produced a plausible reconstruction of *Wiwaxia* that persuaded many researchers this creature belonged to a completely novel phylum. Yet when Nicholas J. Butterfield, then a graduate student at Harvard, looked at *Wiwaxia* in 1990, he suspected it was just a relative of a modern scaleworm called a sea mouse. After some searching, he found evidence that *Wiwaxia* does indeed belong to the phylum Annelida: *Wiwaxia* specimens bear the flattened chitinous hooks that are characteristic of one subclass of living annelid, the polychaetes.

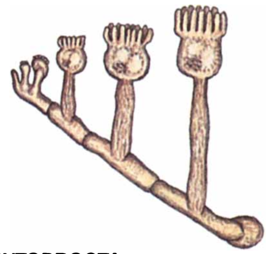
The *Wiwaxia* story has come full circle since Walcott's original and apparently correct conclusion that *Wiwaxia* was an annelid. Even the most peculiar of all the Burgess Shale fossils, whimsically named *Hallucigenia,* has recently been shown by L. Ramsköld of the Swedish Museum of Natural History and Hou Xianguang of the Nanjing Institute of Geology and Paleontology to be in all probability a velvet worm of the phylum Onychophora.

To take another example, for many years the Cambrian echinoderms appeared to be scattered among many taxonomic classes, all of which seemed to spring up at once and without obvious interrelations. More recent analy-

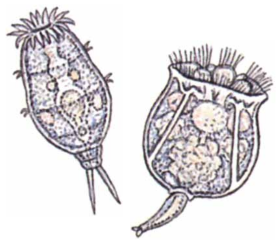
© 1992 SCIENTIFIC AMERICAN, INC



ECTOPROCTA
BRYOZOANS



ENTOPROCTA
BARENTSIA AND OTHERS



ROTIFERA
ROTIFERS

species, reflects variation on these basic physical designs. All the animal phyla that have hard parts arose during the Cambrian. The total number of animal phyla is 26; those pictured here represent only some of the most familiar groups.

ses, however, show a standard evolutionary tree, rather than what appeared to be an evolutionary lawn consisting of many roots back to unknown common ancestors.

Those stories point to a serious problem with all arguments about evolution that rely on taxonomic classification. Some of the fossils that suggest the existence of unique classes are very poor scraps from the geologic table. To establish that a fossil is something truly new, a paleontologist must argue the difference from a lack of features shared by other groups. To establish that a fossil belongs to a known group, one must find the diagnostic traits that prove the relation. When the fragmentary fossils of an unknown organism are first found, they often lack such traits, and on this basis it is easier for paleontologists to assign them to new groups. (Paleontologists, being human, also consider it more fun to find something new.) Because so many of the fossils are fragmentary and potentially subject to reinterpretation, I believe Gould and some other paleontologists have exaggerated the diversity in the Cambrian.

Nevertheless, a Cambrian explosion in animal diversity certainly did occur. Evolutionary biologists are still trying to determine why no new body plans have appeared during the past half a billion years.

James W. Valentine of the University of California at Berkeley has suggested that new variants could have appeared and evolved more rapidly during the early history of life because there was more "open space" in the form of unfilled ecological niches. I do not believe this can be the whole story. David M. Raup of the University of Chicago has estimated that when the biggest of all mass extinctions occurred at the end of the Permian period 230 million years ago, as much as 96 percent of all marine species disappeared. Yet, contrary to Valentine's hypothesis, there is no evidence that any fundamentally new body plans and phyla arose to fill the vacated niches.

It is also hard to accept that the occupation of all niches would completely preclude evolutionary novelty. Today many different body plans exploit the same resources: snails, worms and members of many other diverse phyla all ingest organic particles in mud for food. We must seek another explanation for the dearth of biological innovations since the Cambrian.

One idea worth entertaining is that evolution occurs more slowly today than it did when the earth was young. If evolution has slowed for unknown, peculiar reasons, then perhaps too little time has elapsed for new body plans to evolve.

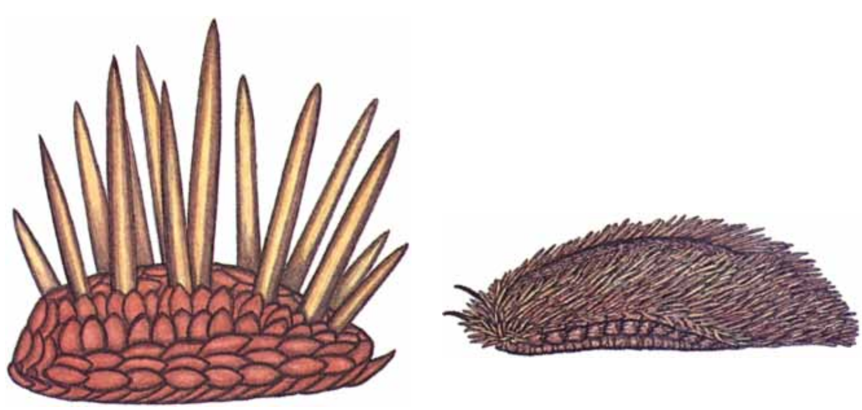
Ten years ago Paul Klerks of the University of Southwestern Louisiana and I decided to test that idea by examining the metal tolerance of invertebrates in Foundry Cove on the Hudson River. Like other coves nearby, Foundry Cove teems with life: water striders whirl about the surface, nabbing unwary midge nymphs; oligochaete worms and insect larvae crowd the muddy bottom and provide food for fish, crabs and prawns. Yet Foundry Cove has a unique distinction: it has perhaps the highest concentrations of toxic cadmium and nickel pollutants in the world.

Lying across the Hudson from Storm King Mountain and West Point, the cove has a venerable military history. During the Revolutionary War, a forge at the site produced chains that were stretched across the Hudson to impede British warships. During the Civil War, the foundry made ammunition. About 40 years ago a military contract brought the manufacture of batteries to the site. Starting in 1953, industrial plants dumped more than 100 tons of nickel-cadmium waste into the cove and nearby river. The dumping continued until complaints by local citizens halted it in the late 1970s.

When Klerks and I first studied the cove in the early 1980s, we found that as much as 25 percent of the bottom sediments consisted of cadmium. Nevertheless, many bottom-dwelling invertebrate species were no less abundant there than in the unpolluted muds of other sites. To learn why, we examined the cadmium tolerance of the most common invertebrate in the cove, an aquatic relative of the earthworm bearing the tongue-twisting name *Limnodrilus hoffmeisteri*.

We found that the *Limnodrilus* from a nearby cove died or showed clear signs of distress when placed in Foundry Cove sediments but that local specimens thrived and reproduced. We raised Foundry Cove worms in clean muds and examined their grandchildren; they too were tolerant of cadmium, which suggested that genes were largely responsible for the tolerance.

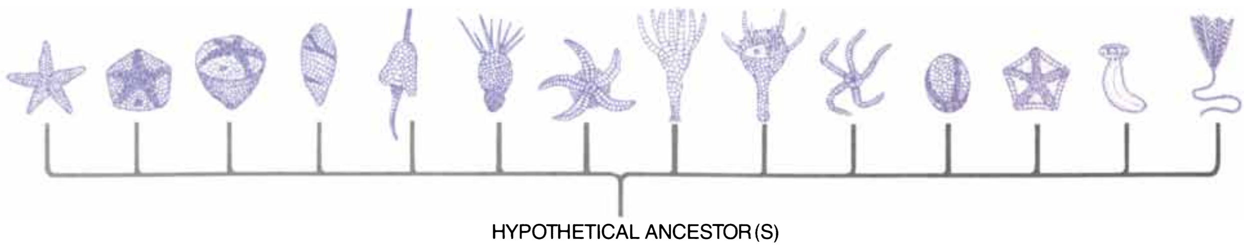
The evolution of cadmium resistance could have taken no more than 30 years. In fact, the genetic variability in nearby populations, together with the high mortality we measured, indicat-

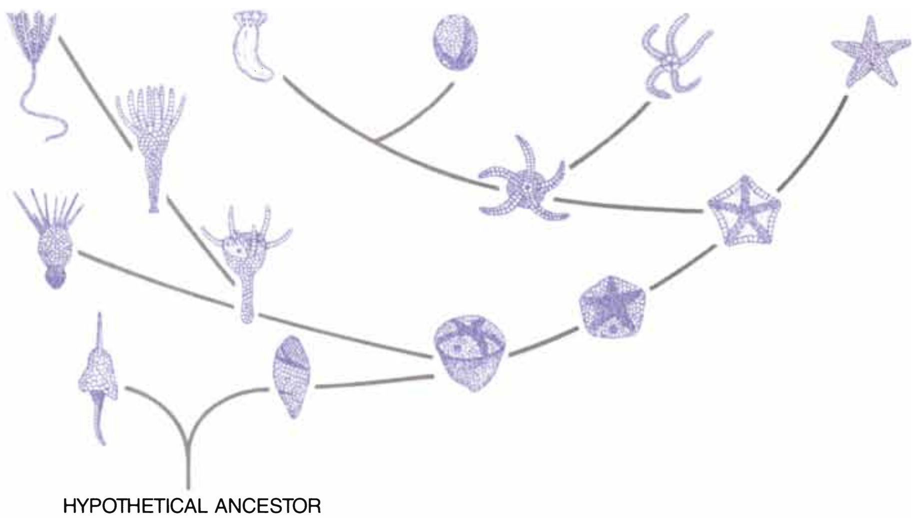


***WIWAXIA* (*left*), a spiny Cambrian fossil, was once believed to represent a phylum unknown in the modern world. Recent work has proved that it is actually related to the sea mouse *Aphrodita* (*right*) of the annelid phylum.**

© 1992 SCIENTIFIC AMERICAN, INC

An Evolutionary Lawn versus an Evolutionary Tree





Differing views of diversity within groups of fossil organisms can sometimes result from preconceptions about the evolutionary relations between the known species. The Cambrian echinoderms were once believed to be a highly diverse group in which many different lineages descended rapidly and separately from a common ancestor, forming a wide evolutionary lawn (*above*). Further study, however, revealed that the same echinoderm fossils could be organized into a more traditional evolutionary tree of branching descent, in which the total diversity is far less (*left*).

ed that the degree of metal tolerance observed could have evolved in just two to four generations—or a couple of years. To prove the conclusion, we exposed worms from an unpolluted site to cadmium-laden sediment and bred the survivors. Sure enough, by the third generation, the descendants had two thirds of the cadmium tolerance found in the Foundry Cove worms.

This capacity for rapid evolutionary change in the face of a novel environmental challenge was startling. No population of worms in nature could ever have faced conditions like the ones humankind created in Foundry Cove. Yet although some species inhabiting nearby waterways are missing from Foundry Cove, most adapted to the unusual conditions.

The rapid evolution of tolerance for high concentrations of toxins seems to be common. Whenever a new pesticide is brought into use, a resistant strain of pest evolves, usually within a few years. The same thing happens to bacteria when new antibiotics are introduced. Luckily for humans, antibiotic resistance seems to be costly for bacteria to maintain, and susceptible strains usually return to dominance when the use of a drug is discontinued temporarily.

One need not turn to poisons to demonstrate the power of natural selection. John A. Endler of the University of California at Santa Barbara has demonstrated with guppies how predators can drive rapid evolution. In the predator-free upper reaches of Trinidadian streams, female guppies often choose mates that have spectacularly large, colorful tails, possibly because those features reflect good health. Bright, showy colors are dangerous, however, where predatory fish abound. Endler tested the effect of predators by breeding guppies in tanks with and without predators. In tanks with many predators, brightly colored males became rare within a few years. In tanks with no predators, they became common.

David N. Reznick and Heather A. Bryga of the University of California at Riverside, working with Endler, showed that natural selection can rapidly alter even an organism's breeding schedule. When the investigators put guppies into a stream lacking predators, the guppies began to reproduce at a later age and devoted more of their food resources to body growth instead of reproduction. When predators were around, natural selection favored guppies that reproduced earlier—before a predator could attack—and spread their reproductive schedules out over many seasons.

Body structures can also evolve rapidly, especially when the formation of new islands or lakes creates fresh ecological spaces that are ripe for invasion. Darwin's finches, a group of closely related species in the Galápagos Islands, probably diverged from a single ancestral species within the past five million years or less. New species with different types of beaks evolved to fill the ecological vacuum, each specializing in a different type of food.

Peter R. Grant of Princeton University and his colleagues were recently able to observe how rapidly natural selection can act on these finches [see "Natural Selection and Darwin's Finches," by Peter R. Grant; SCIENTIFIC AMERICAN, October 1991]. An intense dry spell killed all the plants except those with large, drought-resistant seeds. Because the finches are mainly seed eat-

© 1992 SCIENTIFIC AMERICAN, INC

ers, their mortality was high. These circumstances favored an increase in the average beak size because birds with large beaks could crack open larger seeds. As Grant observed, fluctuations between dry and wet conditions caused continuous bouts of evolution, often in the span of just a few months.

These studies of living groups testify to the vigor and pace of evolution today. J. John Sepkoski, Jr., of the University of Chicago has conducted a fossil survey that provides further support. Sepkoski undertook the titanic job of summarizing the fossil record and cataloguing the diversity of fossil groups over time. His estimate of low taxonomic ranks such as genera probably offers a good indication of the number of species during various periods. He concludes that there have been periods during which the total number of species seems to have been stable and a time at the end of the Paleozoic era when this number dropped cataclysmically. Overall, however, the total number of species seems to have been increasing steadily during the past 60 million years. Clearly, animals are not inhibited from assuming new basic forms by an inability to speciate.

All the evidence from living groups of organisms therefore suggests that contemporary evolution proceeds as fast as ever. Yet if one looks at the fossil record, the pace of evolutionary change can seem quite astonishingly slow.

If you walk along the beach at Scientists Cliffs on the shores of Chesapeake Bay in Maryland, you will come across low prominences of hardened sand containing thousands of fossils from creatures that lived in a shallow sea several million years ago. It is about the easiest fossil collecting imaginable (if you ignore the summer heat and the stinging nettles in the water).

Among the many treasures in the cliffs are Miocene scallop shells named Chesapecten for the Chesapeake Bay tidewater region. Chesapecten scallops were the first North American fossils ever described, in 1687. The earliest members of the group date back to the middle of the Miocene epoch, about 14 million years ago; the Chesapecten scallops have been extinct for roughly three million years. The chain of ancestors and descendants in the strata is nearly unbroken. Joan Miyazaki of the State University of New York at Stony Brook has traced an evolutionary unfolding in these fossil beds that is both grand and majestically slow.

The juveniles of modern scallop species usually have a triangular shape marked by a deep notch through which an "ear" of threadlike material anchors the creature to its substrate. As the scallop matures, its silhouette becomes more circular and the ear relatively smaller. During most of the 11-million-year history of the Chesapecten lineage, the adult form followed the same path in its evolution, graduating from a creature that looked like modern juveniles to modern adults. Miyazaki speculates that the ancient sea in which the Chesapectens lived became gradually deeper, favoring scallops that spent more time living free rather than anchored to a substrate. For whatever reason, the Chesapecten's evolution was surprisingly slow.

Comparable evolutionary transformations in living mollusks can be much more rapid. Contrast the pace of evolutionary changes in the Chesapectens, for example, with those in the dog whelk, *Nucella lapillus,* and the periwinkle *Littorina obtusata*. These mollusks became prey for the European shore crab when it was accidentally introduced into bays in Maine, probably during the early part of this century. Within only a few decades, the dog whelk and the periwinkle evolved thicker, stronger shells that were better able to resist crab attacks.

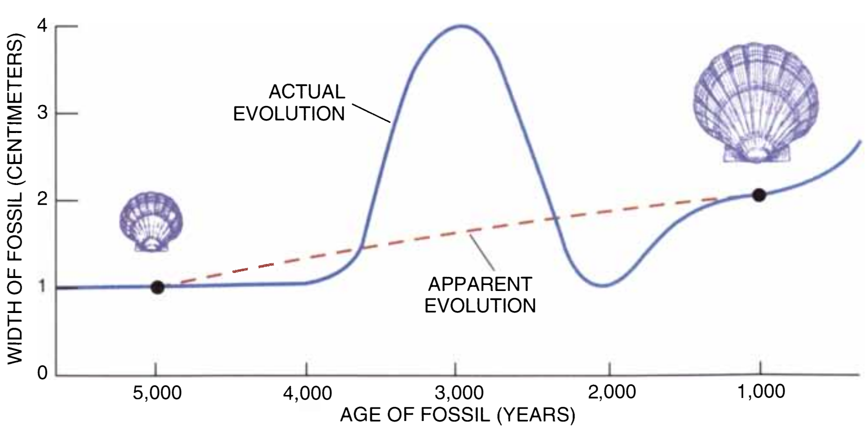
The late George Gaylord Simpson of the American Museum of Natural History, surely the greatest paleontologist of the 20th century, found a similarly slow rate of evolution when he looked at fossil mammals. Modern opossums are only slightly different from their 65-million-year-old Cretaceous ancestors. After extrapolating the rate of opossum transformation, Simpson argued that the evolution of mammals from a reptilian ancestor "can hardly have taken less than 600 million years...which is certainly absurd."

The brilliant evolutionist J.B.S. Haldane provided perhaps the clearest statement of the problem. He scrutinized Simpson's study of fossil horses and found that the height of a tooth feature increased at the startlingly slow rate of about 3.6 percent per million years. Haldane concluded that the evolution of the horses' teeth was so slow that the influence of natural selection appeared to be hardly distinguishable from that of random genetic drift. I have named this Haldane's paradox: How can evolution in living populations be so fast when evolution in the fossil record appears so slow?

The paradox resolves itself to a degree when we recall that paleontologists calculate most evolutionary rates over hundreds of thousands or millions of years. That time scale can bias the detectable rates of change. Suppose you were to measure the water level along a shoreline on January 1 in two successive years. Even if you arrived at low tide in one year and high tide in the next, the measured rate of change would be low—say, one meter per year. On the other hand, if you took a measurement roughly every six hours, the rate of change would appear much higher—one meter every six hours, or 1,460 meters per year.

Unless a change is constant in rate and uniform in direction, the time scale over which it is measured becomes important. Evolutionary rates measured over geologic stretches of time may appear unnaturally slow because the long periods include times of no change, as well as times of rapid change with frequent reversals.

Thus, when geologist Peter M. Sadler of the University of California at Riverside measured the deposition rate of



APPARENT RATES OF EVOLUTION are sensitive to the time scales on which they are measured, which could explain why evolutionary changes seem so slow in the fossil record. In this hypothetical example, the rate of evolution for a fossil's width seems slow (*red*) because periods in which no change occurs and those in which rapid fluctuations occur are lumped together.

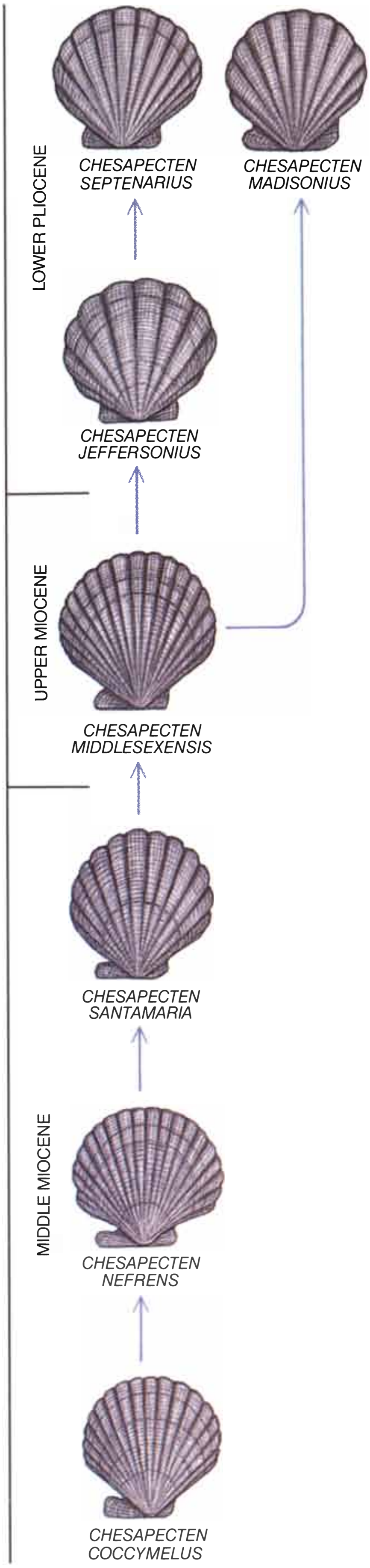
© 1992 SCIENTIFIC AMERICAN, INC

marine sediments at various periods in the geologic record, he found that the rate of sediment accumulation appeared to be lower if it was measured over longer intervals. Paleontologist Philip D. Gingerich of the University of Michigan found the same inverse relation between time scales and the apparent rates of evolution. When he looked for changes in fossil and living species over short intervals, the rates were very high; over longer intervals, the rates appeared lower.

I and many other paleontologists and evolutionary biologists believe that periods of fast and furious evolutionary change alternate with reversals and long periods of little change. The periods of rapid change tend to be lost between the cracks of the paleontologist's time scale. For instance, evolutionary biologist Michael Lynch of the University of Oregon has recently shown that the apparently slow evolution of mammals probably results from natural selection for stable intermediate forms of animals viewed over millions of years.

In short, there is no reason to think that the rate of evolution was ever slower or faster than it is now. Yet that conclusion still leaves unanswered the paradox posed by the Cambrian explosion and the mysterious persistence of those ancient body plans. I have argued that at least part of the answer may depend on the evolution of commitment to a developmental program.

Characteristics that might be regarded as essential to the definition of a group's body plan can sometimes be indistinct or even rather different in the ancestors of that group. C.R.C. Paul of the University of Liverpool and Andrew B. Smith of the Natural History Museum in London have shown that starfish did not just spring from the Cambrian mists into a final form that was never to be altered again. Rather the starfish morphology we know crystallized during the Cambrian over many millions of years. Even the fivefold symmetry of all living echinoderms was not rigidly established at the origin of that group: it was presaged by threefold symmetry in their ancestors. Natural selection slowly sculpted general forms for the echinoderm classes from their shells and soft parts.



CHESAPECTEN FOSSILS show steady trends in evolutionary changes that persisted for more than 10 million years. The shapes of these scallops gradually became more rounded, and the notch marking their point of attachment to the substrate became less pronounced. These changes parallel similar alterations that occur as juvenile scallops mature. In one lineage, these trends eventually stopped, but in the other they continued until the scallops became extinct.

In short, the evidence currently suggests that the defining characters of a group gradually congeal into a more immutable condition after a long period of early evolutionary plasticity. In response to natural selection pressures, developmental programs may evolve to restrict the degree of change in successful body plans. We can only speculate about what genetic mechanisms might permanently set development, but more and more genes are being shown to have similar controls on the early developmental patterns of distantly related species. The Cambrian may have been a period in which the genetic programs that control embryonic body plans locked into the forms we now recognize.

The argument for developmental constraint has considerable strengths. Many biologists have reasoned that because development is an exquisitely fine-tuned process, it cannot be changed radically without difficulty: mutants bearing developmental aberrations are usually defective and die quickly. The theory can explain both the diversification of forms in the Cambrian and the subsequent failure of new body plans to arise after the late Permian extinctions. It also explains why the rate of speciation, as measured by Sepkoski, continues to be high: the changes that correspond to differences between closely related species are not constrained developmentally.

Nevertheless, I must admit to some doubts. Can developmental constraints really account for the stability of body plans over several hundred million years? Can it be true that echinoderms have remained echinoderms because their development cannot be disrupted by natural selection?

The constraints cannot be absolute, because developmental rules are sometimes broken. Tree frogs provide a spectacular example [see "Marsupial Frogs," by Eugenia M. del Pino; SCIENTIFIC AMERICAN, May 1989]. All adult frogs have about the same shape. A few species of tree frog, however, have special pouches into which they lay unusually large, yolky eggs. The early embryology of these species differs from that of other tree frogs to allow the embryo to develop astride a large egg on dry land, rather than in the water.

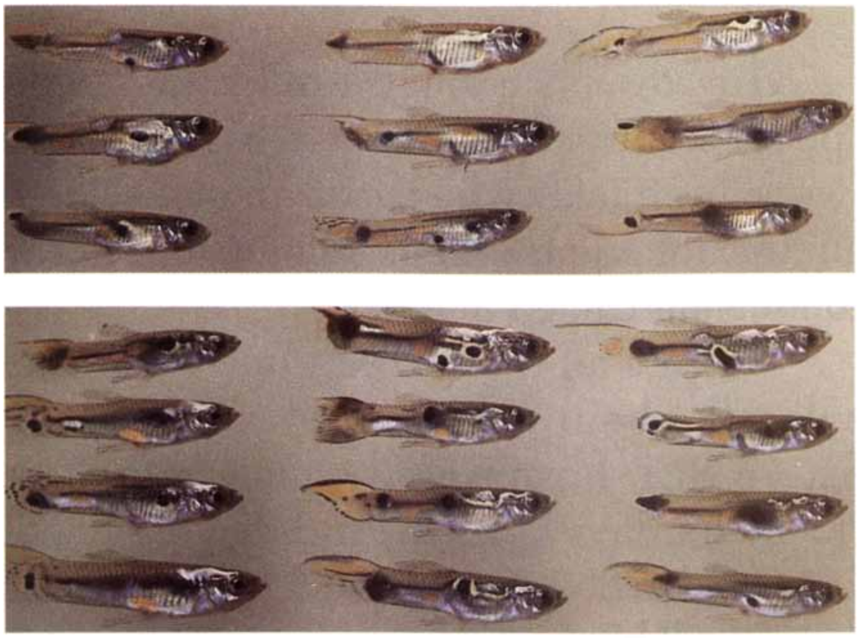
This alteration affects gastrulation, the most crucial period in development, when the future identities of tissues are determined. As the embryologist Lewis Wolpert once said, "It is not

© 1992 SCIENTIFIC AMERICAN, INC

birth, marriage, or death, but gastrulation which is truly the most important time in your life." The fact that development can be altered drastically indicates that the early embryo is not always impervious to important developmental alterations.

The same message comes from recent research on the development of echinoderms. Gregory Wray of Vanderbilt University and Rudolph A. Raff of Indiana University have shown that the embryos of sea urchins are fantastically diverse. The larvae of closely related species sometimes have radically different forms—some are adapted for a long life of swimming and feeding on plankton, whereas others are suited for a short nonfeeding period while they are dispersed by currents. These ecological specializations entail monumental differences in the developmental patterns of the larvae and even in the parts of the embryo used to form adult structures. Yet the adults of these species are virtually indistinguishable: they all have an ovoid skeleton with spines, move on tubular "feet" and scrape rocks with their jaw.

Such diversity indicates that, contrary to the argument for developmental constraint, the adult form must have remained constant because it worked well, not because it was incapable of changing. Although development clearly guides the hand of morphological evolution, natural selection may be the major influence maintaining the sets of features that characterize the phyla.



GUPPIES can evolve significantly within just a few generations to meet changing environmental conditions. When predators are common, natural selection favors guppies that are dull and transparent (*top*). When predators are few, male guppies tend to have larger, more brightly colored tails for attracting mates (*bottom*).

The Cambrian explosion thus remains a mystery. The survival of the body plans that arose during that period seems likely to tell us something important about the patterns of evolution. Working from the assumption that Cambrian life was extraordinarily diverse, Gould has suggested that chance—not natural selection—played the larger role in determining which evolutionary lineages survived and which became extinct. Yet the Cambrian fauna may have been less diverse than he and others have assumed. Derek E. G. Briggs and Matthew A. Wills of the University of Bristol, working with Richard A. Fortey of the Natural History Museum in London, have found signs to that effect: for example, they have demonstrated that modern arthropods seem no less diverse in body form than do the arthropods of the Cambrian.

Moreover, as important as the evolution of the body plans was, the transformation of them since the Cambrian has been extensive. One cannot lightly dismiss the possibility that modern animals do represent progress beyond or improvement on their ancient forebears. Darwin, from the beginning, criticized conceptions of evolutionary progress that would place human beings at the apex of a ladder of life. The progressive trends seen within a group do not imply the resulting organisms' superiority to anything other than their own antecedents. Yet many evolutionary lineages do show recognizable responses to the appearance of predation, variations in climate and water depth as well as other changes.

I argue that these evolutionary trends represent classical Darwinian progressive improvements in response to an environmental challenge. The incremental evolution of mammals from mammallike reptiles, for example, took more than 100 million years to complete and shows progress toward better functioning in a terrestrial environment. The evolution of a secondary palate increased the efficiency of mastication. The teeth changed from simple reptilian cones that were repeatedly replaced during a lifetime to more complex shapes that were replaced only once. Even the jaw joint changed from one fulcrum to another; in some transitional fossils the reptilian jaw joint coexists with the mammalian jaw joint.

If one compared modern reptiles and mammals, it would seem impossible for the articulation of the jaw to change from one set of bones to another without a monstrous, drastic (and highly unlikely) mutation, but the fossils prove that the mammalian jaw underwent gradual evolution through intermediate forms. Far from arguing against the importance of natural selection, as Gould has implied, the fossil record testifies to its pervasive influence.

Are the phylum-level body plans, so ancient and so durable, truly the optimal solutions to the problems of survival and reproduction, reached through an early, fast bout of natural selection before development congealed? Or are they just random combinations of characters assembled by accidents of history? I think the best to be said for now is that there is some truth in both alternatives. Evolution at the species level continues unabated, but variation in the surviving body plans does not seem to occur. For whatever unknown reasons, there will probably never again be an explosion of animal diversity on the earth like the one that took place sometime around the early Cambrian.

FURTHER READING

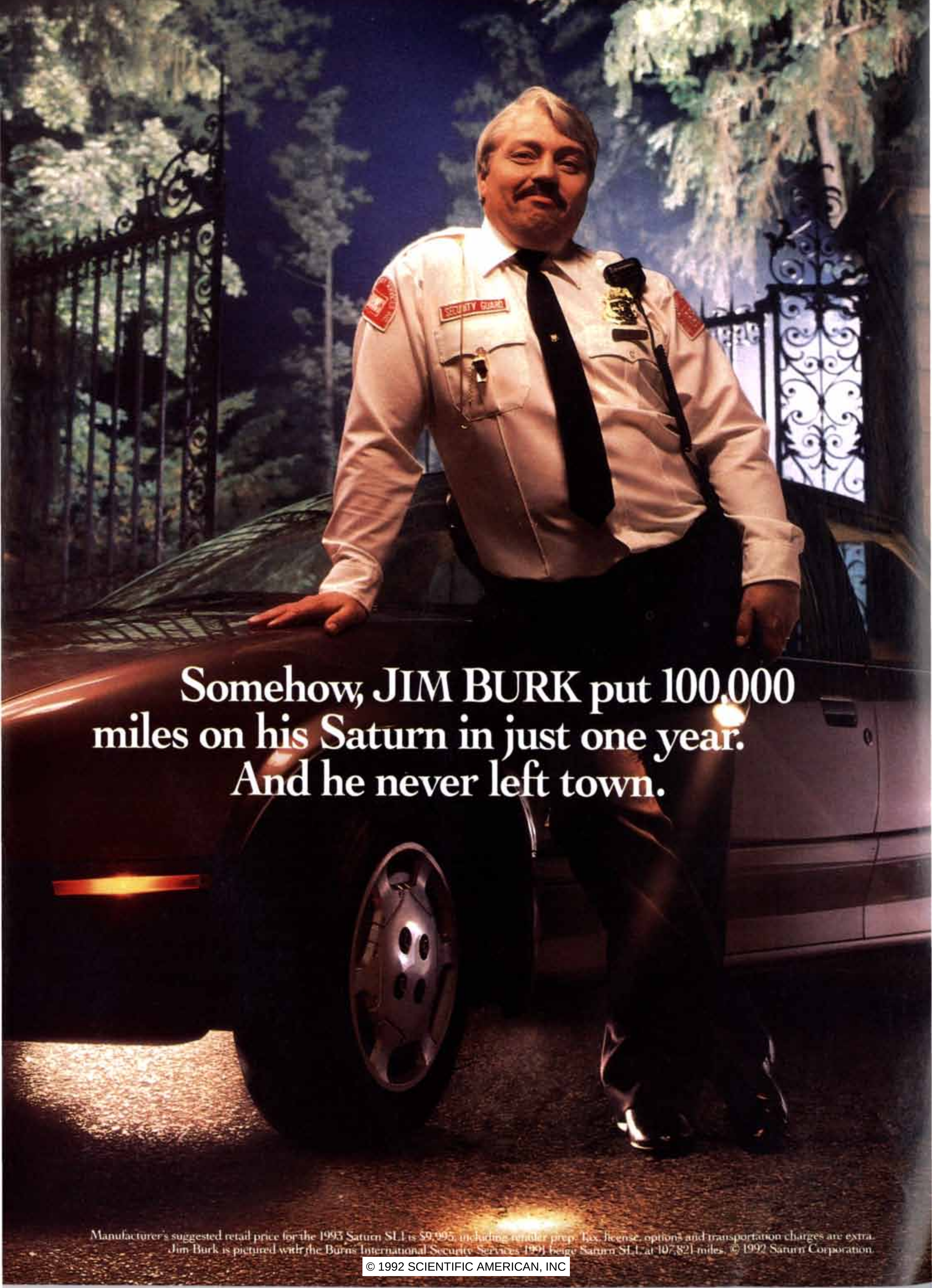
GENETICS, PALEONTOLOGY, AND MACROEVOLUTION. J. S. Levinton. Cambridge University Press, 1988.

THE EMERGENCE OF ANIMALS: THE CAMBRIAN BREAKTHROUGH. M.A.S. McMenamin and D.L.S. McMenamin. Columbia University Press, 1990.

METAZOAN PHYLOGENY AND THE CAMBRIAN RADIATION. D. H. Erwin in *Trends in Ecology and Evolution*, Vol. 6, No. 4, pages 131–134; April 1991.

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC



© 1992 SCIENTIFIC AMERICAN, INC

Linguistic Origins of Native Americans

Scholars have long wondered how Native Americans settled the New World. Recent research indicates that their many tongues belong to just three families, implying three waves of immigration from Asia

by Joseph H. Greenberg and Merritt Ruhlen

A little over two centuries ago Sir William Jones, an English jurist serving in India, was struck by remarkable similarities among Sanskrit, Classical Greek and Latin. He proposed that these languages, and probably Gothic and Celtic as well, had "sprung from some common source, which, perhaps, no longer exists." This source became known, in the following century, as Proto-Indo-European—a protolanguage that linguists have since labored to reconstruct [see "The Origins of Indo-European Languages," by Colin Renfrew; SCIENTIFIC AMERICAN, October 1989, and "The Early History of Indo-European Languages," by Thomas V. Gamkrelidze and V. V. Ivanov; SCIENTIFIC AMERICAN, March 1990].

Jones, however, did not reconstruct a syllable. He reached his conclusions by observing, as he put it, "a stronger affinity, both in the roots of verbs, and in the forms of grammar, than could possibly have been produced by accident" [*see bottom illustration on page 96*]. This evolutionary hypothesis was not lost on scholars interested in New World languages. In 1789, only three years after Jones's celebrated discourse, Thomas Jefferson wrote: "I endeavor to collect all the vocabularies I can, of American Indians, as of those of Asia, persuaded, that if they ever had a common parentage, it will appear in their languages."

Yet although 19th-century scholars identified hundreds of American languages and grouped them into families, none of them ventured the more comprehensive taxonomy that Jefferson had envisaged. The traditional account instead multiplied families, until the number reached about 60 in North America and about 100 in South America, far greater than the number in the Old World, where, for example, Africa has but four.

These estimates are puzzling because taxonomic diversity normally increases with time. Yet most archaeologists have long agreed that human settlement in the Old World substantially predates that in the New. The current consensus is that modern humans emerged at least 100,000 years ago, probably in Africa, and did not reach the Americas until about 12,000 to 20,000 years ago. How could the American languages have diversified to such a great extent?

The difficulty called for a more comprehensive classification. But in the early years of this century, when Alfred L. Kroeber and Edward Sapir first attempted to reduce the many American languages to a handful of larger families, they met with vigorous opposition from such anthropologists as Franz Boas, Pliny Goddard and Truman Michelson. These opponents did not seriously doubt that there were similarities among the American language groups. What they disputed—and what many dispute even today—was the origin of these similarities. Whereas Kroeber and Sapir insisted that the similarities stemmed from a common heritage, and were thus genetic in nature, Boas and his followers attributed the similarities to the diffusion of words from one language family to another.

Our research confirms the genetic approach. By comparing the most conservative elements in the vocabularies of hundreds of languages of North and South America, one of us (Greenberg) found just three families. Because each family shows closer affiliation with Asian families than with the other two American groupings, the tripartite division implies there were at least three distinct migrations from Asia. This hypothesis finds confirmation in the research of physical anthropologists.

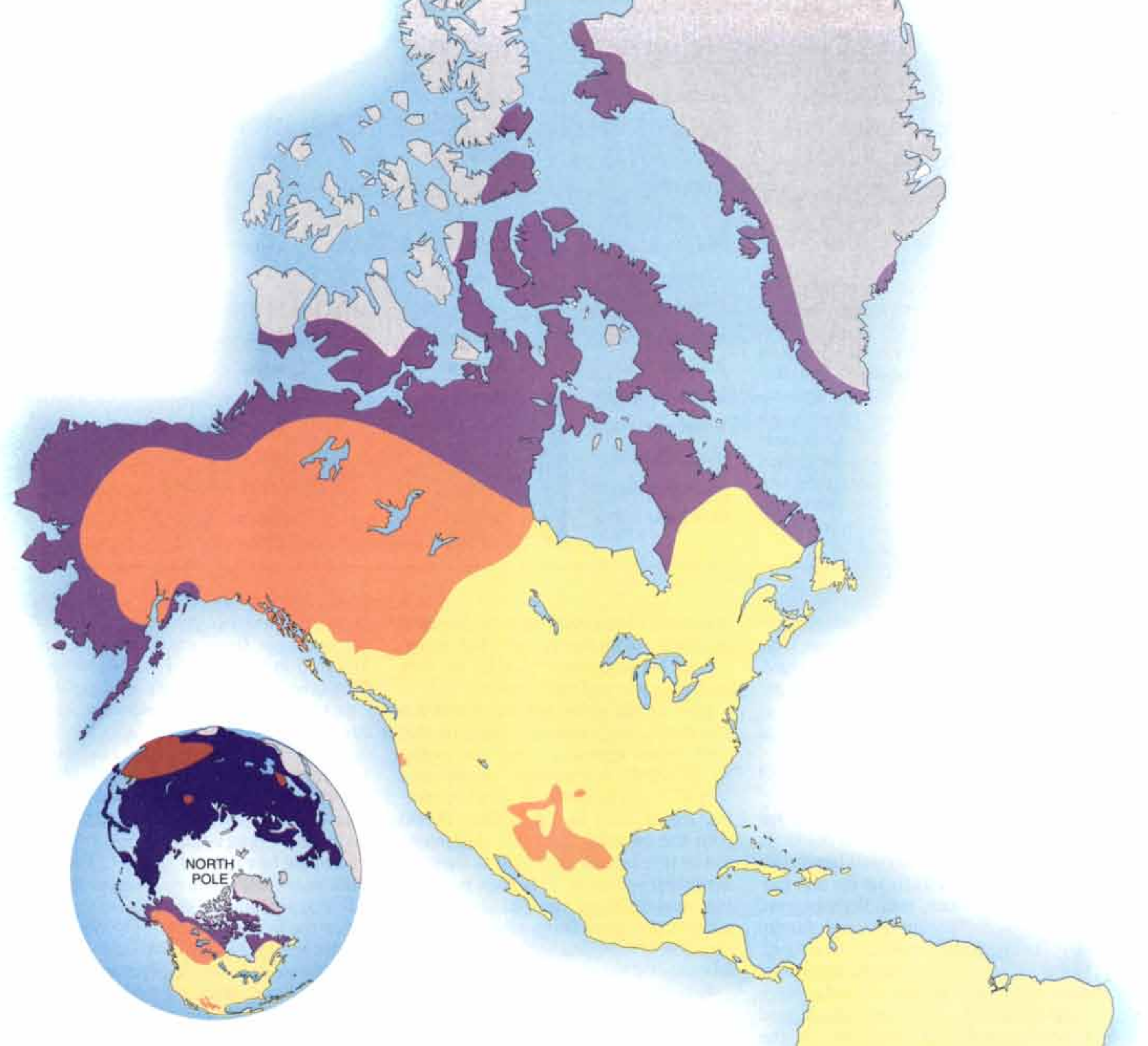
The traditionalists who oppose our classification do not offer a better one. Instead they assert that by comparing languages two at a time and in great depth they will arrive at the true system—in another 50 to 100 years. We believe such work is misguided. To systematize a jumble of languages—or rocks, or animals—one must compare them as a group. Moreover, the multilateral approach has worked before. When Greenberg used it to classify the African languages some 40 years ago, traditionalists in that field opposed the method. Today everyone—even the traditionalists—embraces its results.

Those who compare languages two by two are simply ignoring much relevant evidence. Scholars related Albanian to English not by making a systematic comparison of the two languages in isolation but by establishing that each belonged to the Indo-European family. Indeed, Indo-Europeanists have never used a binary approach.

Our system of multilateral analysis uncovers precisely those relations that tend to escape notice in the binary ap-

JOSEPH H. GREENBERG and MERRITT RUHLEN collaborate in the comparative study of languages. Greenberg is professor emeritus of anthropology and linguistics at Stanford University. His chief interests are the universal features of language and the historical classification of language. He is past president of the Linguistic Society of America and has been elected a member of the American Philosophical Society, the American Academy of Arts and Sciences and the National Academy of Sciences. Ruhlen is an independent researcher based in Palo Alto, Calif. He earned a doctorate in linguistics from Stanford in 1973 and worked on the Language Universals Project. He is author of *A Guide to the World's Languages.*

© 1992 SCIENTIFIC AMERICAN, INC



GREENBERG'S CLASSIFICATION places America's many languages into just three families. Eskimo-Aleut (*purple*) and Na-Dene (*orange*) belong to the Old World groups known as Eurasiatic and Dene-Caucasian, respectively (*inset*). Amerind (*yellow*) is related to Eurasiatic. Amerind was the first family to enter the New World, Eskimo-Aleut, the last.

proach. We compare hundreds of languages at a time—a search in breadth rather than an analysis in depth—by examining a list of several hundred words. This list contains words that denote universal concepts, such as personal pronouns, body parts and aspects of nature (water and fire, for example). Because such concepts are rarely borrowed, languages seldom have occasion to borrow their names. English provides an illustration of this rule. Although it has borrowed many words from many languages, most of its basic vocabulary derives from Proto-Germanic. English has "one, two, three, I, mine, father, water"; German has "*ein, zwei, drei, ich, mein, Vater, Wasser*."

A comparison of the basic vocabularies of hundreds of languages from North and South America led Greenberg to group the many postulated families into just three: Eskimo-Aleut, Na-Dene and Amerind. The first two—Eskimo-Aleut in the Arctic and Na-Dene in Canada and the southwestern U.S.—had long been accepted, and so the innovation consisted in grouping all the other American languages under Amerind. It contains 11 subfamilies, distributed throughout much of North America and all of South America [*see illustration on this page*].

In support of Amerind, Greenberg

© 1992 SCIENTIFIC AMERICAN, INC

Defining a Family by a Single Linguistic Innovation: T'ANA

Evidence in its daughter languages implies that Proto-Amerind had a root that sounded like T'ANA, meant "child" and assumed three vocalizations that indicated gender. Because the etymology runs through all of Amerind's 11 branches but is not found in any other group, it ties the family together and distinguishes it from others. Branches appear in the first column. Almosan-Keresiouan and Chibchan-Paezan are divided, and each thus occupies two rows. All daughter languages are modern save Proto-Uto-Aztecan, which is reconstructed.

LANGUAGE FAMILY	LANGUAGE	FORM	MEANING
AMERIND	PROTO-AMERIND	**T'A'NA**	"CHILD, SIBLING
Almosan	Nootka	**t'an'a**	"child"
Keresiouan	Yuchi	**tane**	"brother"
Penutian	Totonac	**t'ána-t**	"grandchild"
Hokan	Coahuilteco	**t'an-pam**	"child"
Central Amerind	Proto-Uto-Aztecan	***tana**	"daughter, son"
Chibchan	Miskito	**tuk-tan**	"child, boy"
Paezan	Warrau	**dani-**	"mother's sister"
Andean	Aymara	**tayna**	"firstborn child"
Macro-Tucanoan	Masaca	**tani-mai**	"younger sister"
Equatorial	Urubu-Kaapor	**ta'ĩn**	"child"
Macro-Carib	Pavishana	**tane**	"my son"
Macro-Panoan	Lengua	**tawin**	"grandchild"
Macro-Ge	Tibagi	**tog-tan**	"girl"

proposed about 300 etymologies, or groups of words that he believes have all evolved from a single ancestral word. The members of each such group are called cognates. Recent work by one of us (Ruhlen) has raised the number of etymologies to about 500.

Some of these roots are distributed so broadly that it is difficult to understand how they were overlooked for so long. The main reason, no doubt, is that specialists in American languages have each tended to focus on one language family. Thus, even if there were similar words running through family after family, nobody would notice them.

A good example is furnished by an Amerind root whose sounds were roughly TANA, TINA or TUNA and whose meaning fell somewhere in the range of "child, son, daughter" (the capital letters signify that the sounds are approximations). No one who carefully compares the vocabulary of Amerind languages from North and South America can fail to be impressed by the very high frequency of such terms.

How should we explain this broad distribution? One possibility might be that such terms appear around the world, as do words resembling "mama" and "papa." Unfortunately for this hypothesis, forms such as TANA and TUNA, with the meaning "son" or "daughter," are as rare outside Amerind as they are abundant within it. This root not only ties Amerind together but also distinguishes Amerind from other language families. It is, as linguists say, an exclusive innovation of the Amerind family.

Recent research by Ruhlen appears to explain why the first vowel of the root varies and why the root finds widespread use in words denoting both the sexes (son/brother and daughter/sister) and the neutral form (child/sibling). The reason is that Proto-Amerind, the original language from which all modern Amerind languages derive, had three forms, or grades, of the root in question, in which the first vowel was correlated with sex as follows: T'ANA "child, sibling," T'INA "son, brother, boy" and T'UNA "daughter, sister, girl." (The apostrophe represents a glottal stop after the "T"—a sound heard in the Cockney pronunciation of "bottle.")

As might be expected, in the 12,000 or more years since Amerind began to divide into subfamilies, the correlation between the initial vowel and the original gender has often been lost. As a result, many forms that are clearly cognates of the others now show the "wrong" vowel. One example of this kind is Proto-Algonquian *tāna "daughter," where the first vowel is *ā rather than *ū. (The asterisk signifies that the form has been reconstructed on the basis of the modern daughter languages.) Most likely this discrepancy is the result either of the first vowel assimilating the timbre of the second vowel or of the *a*-form of the root being extended, by analogy, throughout the language at the expense of the *i*- and *u*-forms. Such analogical extension is common in linguistic history. In English, for instance, the *-ed* form of the past tense of regular verbs (as in "kick/kicked") is extended by some speakers to the past tense of irregular verbs (as in "see/see'd").

It is noteworthy that the vowels *i* and *u* proposed for these masculine and feminine kinship terms coincide with the gender system in two major Amerind subgroups of South America and also in the Chinook language of Oregon. These agreements are too numerous to

	SANSKRIT	CLASSICAL GREEK	LATIN	OLD IRISH	GOTHIC
I carry	bhár-āmi	phér-ō	fer-ō	bir-u	baír-a
thou carriest	bhár-asi	phér-eis	fer-s	bir-i	baír-is
he carries	bhár-ati	phér-ei	fer-t	ber-id	baír-ith
we carry	bhár-āmas	phér-omen	fer-imus	ber-mi	baír-am
you carry	bhár-atha	phér-ete	fer-tis	ber-the	baír-ith
they carry	bhár-anti	phér-ousi	fer-unt	ber-it	baír-and

VERBAL VESTIGES of a common ancestor led William Jones, an 18th-century English jurist, to place these five ancient languages in one family, now called Indo-European. English is most closely related to Gothic.

© 1992 SCIENTIFIC AMERICAN, INC

LANGUAGE	FORM	MEANING	LANGUAGE	FORM	MEANING
PROTO-AMERIND	**T'I'NA**	"SON, BROTHER, BOY"	PROTO-AMERIND	**T'U'NA**	"DAUGHTER, SISTER, GIRL"
Yurok	**t^sin**	"young man"	Coeur d'Alene	**tune**	"niece"
Mohawk	**-'tsin**	"male, boy"	Yuchi	**$t^{s'}$one**	"daughter, son"
Molale	**pnē-t'in**	"my elder brother"	Central Sierra Miwok	**tūne-**	"daughter"
Yana	**t'inī-si**	"child, son, daughter"	Salinan	**a-t'on**	"younger sister"
Cuicatec	**'díínó**	"brother"	Taos	**-t'út'ina**	"older sister"
Changuena	**sin**	"brother"	Lenca	**tuntu-rusko**	"younger sister"
Millcayac	**tzhœng**	"son"	Cayapa	**t^suh-ki**	"sister"
Tehuelche	**den**	"brother"	Tehuelche	**thaun**	"sister"
Tiquie	**ten**	"son"	Tiquie	**ton**	"daughter"
Mocochi	**tin-gwa**	"son, boy"	Morotoko	**a-tune-sas**	"girl"
Yagua	**dēnu**	"male child"	Nonuya	**-tona**	"sister"
Tacana	**u-tse-kwa**	"grandchild"	Tacana	**-tóna**	"younger sister"
Guato	**china**	"older brother"	Piokobyé	**a-ton-kä**	"younger sister"

be accidental and too widespread to reflect linguistic borrowing. Indeed, many of them fall on either side of clear geographic discontinuities.

Just as Jones was impressed by the conjunction of roots and affixes, so too do we find in Amerind an equally impressive conjunction of the root in question and various grammatical affixes. Those that may modify the root T'ANA include the pronomial prefixes *na-* "my" and *ma-* "your," both of which appear in all 11 Amerind subgroups. The former appears in forms such as Proto-Algonquian **ne-tāna* "my daughter," Kiowa *nō-tō*^{n} "my brother," Paez *ne-tson* "my brother-in-law" and Manao *no-tany* "my son." Such pronomial affixes are among the most stable elements in language: they are almost never borrowed. That entire systems of them could have been systematically transmitted from one language to the next, from British Columbia to Tierra del Fuego, defies the imagination.

Amerind suffixes include diminutive forms that one naturally associates with words denoting children. The Proto-Amerind diminutive **-i'sa* is found in Proto-Algonquian **ne-tān-ehsa* "my daughter," Mixtec *tá'nù i'sá* "younger sister," Esmeralda *tini-usa* "daughter," and Suhin *tino-ice* "young woman." The Proto-Amerind diminutive **-mai* is seen in Luiseño *tu'-mai* "woman's daughter's child," Masaca *tani-mai* "younger sister" and Chapacura *tana-muy* "daughter."

Proto-Amerind deployed an intricate system of suffixes. One such suffix, **-ki*, indicated a reciprocal relation, such as that which makes a single word mean either a man's sister's son or a boy's mother's brother. This suffix—in conjunction with various roots to which it attaches—has been reconstructed for Proto-Siouan as **-thã-ki* "man's sister" and is seen in such modern languages as Pawnee *t'i-'i* "boy, son," Southern Pomo *t'i-ki* "younger sibling," Mazahua *t'i-'i* "boy," Amaguaje *-tsen-ke* "son" and Aponegicran *-thon-ghi* "sister."

The threefold classification of languages implies that no more than three Asian migrations left linguistic traces. Fewer migrations are possible if they gave rise to communities that split on the eastern side of the Bering Strait. To decide on the precise number, one must compare the language families of America and Asia.

Recent work by Russian and American linguists indicates that there probably were exactly three migrations. Eskimo-Aleut is the easternmost member of a vast family that we call Eurasiatic and that Russian scholars call Nostratic. (The two classifications differ slightly. Eurasiatic includes Indo-European, Uralic-Yukaghir, Turkic, Mongolian, Tungus, Korean, Japanese, Ainu, Gilyak, Chukchi-Kamchatkan and Eskimo-Aleut. Nostratic is broader, including also the Dravidian family of southern India, the Kartvelian family of the Caucasus and the Afro-Asiatic family of North Africa and the Middle East.)

Na-Dene's relatives in Asia were recently identified by Sergei Starostin of the Institute of Oriental Studies, Sergei Nikolaev of the Institute of Slavic Studies in Moscow and John Bengtson, an independent linguist in Minneapolis. Starostin began by connecting three Old World families that had hitherto been considered independent: Caucasian, Sino-Tibetan and Yeniseian (a family of central Siberia that has only a single surviving language). Nikolaev then showed that Na-Dene was unmistakably related to Caucasian (which he and Starostin had together reconstructed) and hence by extension to Sino-Tibetan and Yeniseian as well.

In a more comprehensive comparison of all relevant families, Bengtson added Basque (an isolated language of northern Spain) and Burushaski (an isolated language of northern Pakistan) to this family, which has come to be called Dene-Caucasian. Na-Dene proves to be the easternmost extension of Dene-Caucasian. Because that family is distinct from Eurasiatic, Na-Dene could not have split from Eskimo-Aleut in the Western Hemisphere. It must have reached the Americas by means of a separate migration.

Over the past few years, we have compared Amerind with the world's other language families and found that it is most closely related to Eurasiatic. The taxonomic relation is quite distant: whereas Eskimo-Aleut is a member of the Eurasiatic family, Amerind is related to Eurasiatic as a whole. That is, its genetic connection reaches much further back in time.

The first migration, known on archaeological grounds to have occurred some time before 12,000 B.P., gave rise to the Amerind family, which occupied most of the New World at the time of Columbus's arrival in 1492. The second migration, somewhat later, gave rise to the Na-Dene family. Finally, perhaps 4,000 to 5,000 years ago, the final migration took place, bringing the ancestors of the Eskimo and Aleut first to

© 1992 SCIENTIFIC AMERICAN, INC

southwestern Alaska and later across the northern perimeter of North America to Greenland.

A single etymology can illustrate both the unity of Amerind and its ties to the Eurasiatic/Nostratic constellation. The Proto-Amerind root MALIQ'A, meaning "swallow, throat," has left its mark in no fewer than eight of the 11 Amerind subfamilies from Canada to the tip of South America [*see illustration below*]. In Canada's Salish subfamily we find Halkomelem *məlqw* "throat." Down the coast in Oregon we find in Tfaltik, an extinct language of the Penutian subfamily, *milq*, which means "swallow." In Yuman, a subdivision of the Hokan subfamily, this root has become the general word for "throat." In Arizona we find Mohave *malʸaqe* "throat," whereas Akwa'ala, in Baja California, has *milqi* "neck." In Panama, Cuna has *murki* "swallow," where the original *l* has apparently changed to *r*, a very common replacement. In the Andean subfamily the Quechua language has *malq'a* "throat"; in the Equatorial subfamily, the Guamo language has *mirko* "drink."

What is the probability that these similar forms arose independently? One can make a rough estimate by holding the meaning within the narrow semantic range "swallow-throat" and making a number of phonological assumptions. Let us begin by assessing the probability that the Halkomelem and Tfaltik forms resemble each other by accident. Disregard the vowels as less stable than consonants and calculate the chances that the three consonants will accidentally match. Next, limit both languages to only the following consonants: *p, t, t', k, k', q, q', s, m, n, l, r, y, w*. Accept only *m* for the first consonant, *l* or *r* for the second position and *k, k', q, q'* for the third consonant.

Under these assumptions, the chances of an accidental match are (1/13)(2/13)(4/13) = 0.0036413291. If we then round this off to 0.004 and calculate the probability for a random similarity among six families, we obtain $(0.004)^5$ = 0.0000000001024, or about one chance in 10 billion. These rough calculations assume an equal probability of all consonant types. Because this assumption does not hold, the figure will actually be somewhat larger, yet it will still be of the same infinitesimal order of magnitude. So much for accidental resemblances.

LANGUAGE FAMILY		LANGUAGE	FORM	MEANING
AFRO-ASIATIC		Proto-Afro-Asiatic	***mlg**	"to suck, breast, udder"
		Arabic	**mlj**	"to suck the breast"
		Old Egyptian	**mndʸ**	"woman's breast, udder"
INDO-EUROPEAN		Proto-Indo-European	***melg-**	"to milk"
		English	**milk**	"to milk, milk"
		Latin	**mulg-ēre**	"to milk"
URALIC		Proto-Finno-Ugric	***mälke**	"breast"
		Saami	**mielga**	"breast"
		Hungarian	**mell**	"breast"
DRAVIDIAN		Tamil	**melku**	"to chew"
		Malayalam	**melluka**	"to chew"
		Kurux	**melkhā**	"throat"
ESKIMO-ALEUT		Central Yupik	**melug-**	"to suck"
AMERIND		Proto-Amerind	***maliq'a**	"to swallow, throat"
	Almosan	Halkomelem	**məlqw**	"throat"
		Kwakwala	**m'lXw-'id**	to chew food for the baby"
		Kutenai	**u'mqolh**	"to swallow"
	Penutian	Chinook	**mlqw-tan**	"cheek"
		Takelma	**mülk'**	"to swallow"
		Tfaltik	**milq**	"to swallow"
		Mixe	**amu'ul**	"to suck"
	Hokan	Mohave	**malʸaqé**	"throat"
		Walapei	**malqi'**	"throat, neck"
		Akwa'ala	**milqi**	"neck"
	Chibchan	Cuna	**murki-**	"to swallow"
	Andean	Quechua	**malq'a**	"throat"
		Aymara	**malʸq'a**	"to swallow, throat"
	Macro-Tucanoan	Iranshe	**moke'i**	"neck"
	Equatorial	Guamo	**mirko**	"to drink"
	Macro-Carib	Surinam	**e'mōkï**	"to swallow"
		Faai	**mekeli**	"nape of the neck"
		Kaliana	**imukulali**	"throat"

OLD WORLD TIES appear in the etymology of the extremely ancient root MALIQ'A, whose meaning was close to "swallow" or "throat." Cognates appear in eight Amerind branches and in more than one language from each of the listed Old World families. The chances that such resemblances could have occurred by accident are vanishingly small.

© 1992 SCIENTIFIC AMERICAN, INC

AMERICAN LANGUAGES AND PEOPLES form two correlated family trees, the one based on etymologies, the other on genes. Amerind speakers include the Maya, who carved the glyph for *ts'apah,* meaning "was set upright," more than 1,000 years ago (*left*). Among the Na-Dene speakers are the Apaches, who were led in the 19th century by Geronimo (*center*). Eskimo-Aleut speakers, including these Inuit from Canada's Northwest Territory (*right*), range from Siberia to Greenland.

Let us turn now to the question of whether this root can be found in the Old World. As we saw earlier in the case of T'ANA "child," there is no guarantee that elements widespread in Amerind will be found outside that family. In this case, however, cognate forms of this root are scattered through the Old World. The original Russian Nostraticists, the late Vladislav Illich-Svitych and Aaron B. Dolgopolsky (now at the University of Haifa), have reconstructed a Nostratic root **mälgi* "to suck the breast, to nurse." This root connects Proto-Afro-Asiatic **mlg* "to suck the breast" (as in the Arabic *mlj*), Proto-Indo-European **melg-* "to milk," as well as the noun "milk" and Proto-Finno-Ugric **mälke* "breast" (as in Saami *mielgâ*). We have found cognate forms in Eskimo-Aleut such as Central Yupik *melug-* "to suck." Finally, the Dravidian family displays apparent cognates in forms such as Kurux *melkhā-* "throat" and Tamil *melku* "to chew."

The range in meaning displayed by these families suggests that the ultimate ancestor of this root meant "to nurse, to suck the breast," a meaning preserved in Afro-Asiatic. In Indo-European there was a slight semantic shift from the notion of nursing to that of milking, whereas Uralic shows a different shift: to the noun "breast." In Dravidian the meaning has shifted to "chew," a natural semantic connection for anyone who has ever watched a baby nursing, and "throat." In Eskimo the meaning has become "to suck" in general, without specific reference to the female breast. Finally, in Amerind this root became the general word for "to swallow" and "throat."

Support for the Amerind hypothesis came from an unexpected quarter in 1988, a little more than a year after it was first announced. A team of geneticists led by L. L. Cavalli-Sforza of the Stanford University School of Medicine discovered that Native Americans fell neatly into three distinct groups whose boundaries essentially coincided with those of their respective language families [see "Genes, Peoples and Languages," by L. L. Cavalli-Sforza; SCIENTIFIC AMERICAN, November 1991]. This independent corroboration virtually confirms the validity of the Amerind family because the probability that the biological and linguistic classifications would coincide fortuitously is very small indeed.

Yet a third line of evidence supporting a tripartite classification of Native Americans has been developed by Christy G. Turner II of Arizona State University. A specialist in human dentition, Turner found that on the basis of their teeth, New World populations fall into the same three groups [see "Teeth and Prehistory in Asia," by Christy G. Turner II; SCIENTIFIC AMERICAN, February 1989]. Finally, in 1990 Douglas C. Wallace of the Emory University School of Medicine reported preliminary results of the analysis of mitochondrial DNA in Native American populations, and this analysis also appears to support the Amerind hypothesis.

We must hasten to add that the close correspondence of biological and linguistic classifications does not mean that genes determine the language one speaks. That depends solely on the community in which one is raised. The classifications correspond because the same processes that lead to linguistic divergence also give rise to genetic divergence. When a group of people depart from their homeland and move, say, to some distant island, they take with them both their language and their genes. From this time on, their language and their gene pool will diverge from those of the group left behind. It is for this reason that the classifications correspond so nicely.

The evidence of comparative linguistics indicates that the Americas were originally settled by three major migrations from Asia. There are, of course, many unresolved problems, such as how the Amerind family initially broke up in its spread through North and South America. But the recent discoveries at least, in part, fulfill Jefferson's hope that one day the languages of Native Americans would illuminate their relations to one another and reveal the Asian origins of the first Americans.

FURTHER READING

THE SETTLEMENT OF THE AMERICAS: A COMPARISON OF LINGUISTIC, DENTAL, AND GENETIC EVIDENCE. Joseph H. Greenberg, Christy G. Turner II and Stephen L. Zegura in *Current Anthropology,* Vol. 27, pages 477–497; 1986.

LANGUAGE IN THE AMERICAS. Joseph H. Greenberg. Stanford University Press, 1987.

THE AMERICAN INDIAN LANGUAGE CONTROVERSY. Joseph H. Greenberg in *Review of Archaeology,* Vol. 11, No. 2, pages 5–14; Fall 1990.

A GUIDE TO THE WORLD'S LANGUAGES, Vol. 1: CLASSIFICATION. Merritt Ruhlen. Stanford University Press, 1991.

EVOLUTION OF HUMAN LANGUAGES. Edited by John Hawkins and Murray Gell-Mann. Addison-Wesley, 1992.

© 1992 SCIENTIFIC AMERICAN, INC

Astronomy in the Age of Columbus

Aided by his wildly erroneous conception of the earth's circumference, Columbus redrew the globe, perhaps encouraging others to realign the heavens

by Owen Gingerich

Two events of astronomical interest took place in 1492. One was the explosion of a brilliant fireball over central Europe, which dropped its stony meteorite near the Alsatian town of Ensisheim. The other was Columbus's discovery of the New World.

The impressionable young Albrecht Dürer witnessed the fireball while en route to Italy and painted the magnificent phenomenon on a small wood panel. He used the other side for an oil of St. Jerome, however, and his painting of the meteoritic explosion stayed hidden from sight for several centuries. It came to light again about two decades ago, when the St. Jerome painting was lent to the Fitzwilliam Museum in Cambridge, England. Meanwhile the Ensisheim meteorite, which was kept in the parish church and later in the Hôtel de Ville at Ensisheim, remained practically unknown. Not until the past few decades has the Ensisheim stone—the oldest precisely dated meteorite in Europe—captured the attention of meteoriticists.

Curiously enough, what seems like the nonastronomical event of 1492—Columbus's voyage across the Atlantic and back—may well have had a genuinely significant impact on astronomical thought. Even though Columbus was wrong in his belief that he could sail westward to China and Japan, his pioneering venture and the voyages that followed vividly demonstrated that ancient knowledge—particularly geographic information—was woefully incomplete. The geographic revolution of the New World paved the way for unorthodox astronomical ideas, including the possibility of a radical, sun-centered cosmology.

OWEN GINGERICH is a senior astronomer at the Harvard-Smithsonian Center for Astrophysics in Cambridge and chairman of the history of science department at Harvard University. This is his fifth article for *Scientific American.* For many years Gingerich crisscrossed Europe and America searching for copies of Copernicus's *De revolutionibus*; he has inspected more than 500 16th-century copies looking for early annotations. Two anthologies of his articles are just being published: *The Great Copernicus Chase and Other Adventures in Astronomical History* (Sky Publishing and Cambridge University Press) and *The Eye of Heaven: Ptolemy, Copernicus, Kepler* (American Institute of Physics).

What was the state of astronomical knowledge when Columbus made his voyage? A good starting point for the answer is Hans Holbein's *The Ambassadors,* painted in 1533, one of the great treasures of the National Gallery in London. Between the ambassador from the French court and his scholarly friend, the bishop of Lavaur, stands a table filled with books and instruments. At first glance these artifacts distribute nicely between the heavens, the earth and the sea, represented, respectively, by the celestial dials and globe on top, the earthly books and lute on the shelf, and a very fishy form near the floor.

More fundamentally, the objects constitute an allegory of the four topics of the advanced university curriculum: astronomy, arithmetic, music and geometry. The instruments represent astronomy, and Peter Apian's *Eyn newe unnd wolgegründte underweysung aller Kauffmanss Rechnung* of 1527, lying open on the shelf, portrays arithmetic. The lute and a songbook open to the Lutheran *Kom Heiliger Geyst* signify music. Geometry is exemplified not only by the challenging perspectives of the mosaic floor (an Italian mosaic in the shrine of Edward the Confessor in Westminster Abbey, unique in England) and the lute—a favorite drawing exercise for Renaissance artists—but also by the curious fishlike object hovering above the floor. Closer inspection, from a vantage point that foreshortens the image, reveals it to be an elongated anamorphic depiction of a human skull, perhaps a pun on the name of the artist (Holbein, "hollow bone").

The skull, a symbol of mortality, brings in yet another level of Renaissance metaphor, reminding us that any quest for earthly knowledge is transitory and ephemeral. The theme is reinforced by the broken lute string, also a traditional symbol of death and decay. In contrast to the highly visible studies of the quadrivium stand the eternal mysteries, symbolized by the crucifix half-concealed behind the curtain at the upper left corner of the painting. The scholarly pursuits may be focused and central, but the larger truths lie hidden beyond mortal powers.

The era of *The Ambassadors* was still a time when some long-gone golden age was thought to hold the keys to the universe, and newness was not yet a virtue. Nevertheless, astronomy took an honored place in the curriculum, for it described the physical arena in which the human drama took place.

The earth—a composite sphere of earth, water, air and fire—was solidly fixed in the middle of the cosmos. Around it were the spheres of the seven planets (counting both the moon and sun) and an eighth sphere containing the fixed stars—fixed with respect to one another but actually spinning at a dizzying rate, once every 24 hours. And beyond that were God the Father with his angels and the elect in a state of eternal bliss. A woodcut in the *Nuremberg Chronicle,* that great coffee-table book of 1493, sets forth the classical cosmology in all its glory.

The *Nuremberg Chronicle* was designed before Europe had heard about

© 1992 SCIENTIFIC AMERICAN, INC

HOLBEIN PAINTING, *The Ambassadors,* contains representations of the university curriculum as of 1533: arithmetic, astronomy, geometry and music. Among the astronomical instruments is a globe depicting parts of the New World.

Columbus, and a single look at it debunks one of the most widely disseminated myths about the Italian navigator's voyage. Since the ancient Greeks, people had known the earth as a sphere. Aristotle had taught that the earth had to be a globe because if bits of heavy terrestrial material were dropped into the center of the universe, they would naturally pile up in the form of a sphere. Almost as an afterthought he added that the shape of the shadow of the earth on the moon at the time of a lunar eclipse demonstrated the correctness of his archetypal idea.

Yet every Columbus Day the story is vividly retold of how Columbus had to persuade Isabella and Ferdinand that the world was round. Had Christendom forgotten the round earth? In truth, it is early 19th-century Americans who were forgetful, and the doctrine they were trying to forget was the standard British view that one of theirs, Sebastian Cabot, was the first to make landfall in North America—Columbus had merely found some small islands in the Indies.

In the aftermath of the Revolution, Americans desperately needed some non-British heroes. Columbus filled the bill even though no one knew much about him until novelist Washington Irving visited Spain, found a rich lode of source materials and produced a widely read biography. Unfortunately, Irving mixed fiction with fact, and one of his most graphic scenes, set in Salamanca, was wildly imaginative.

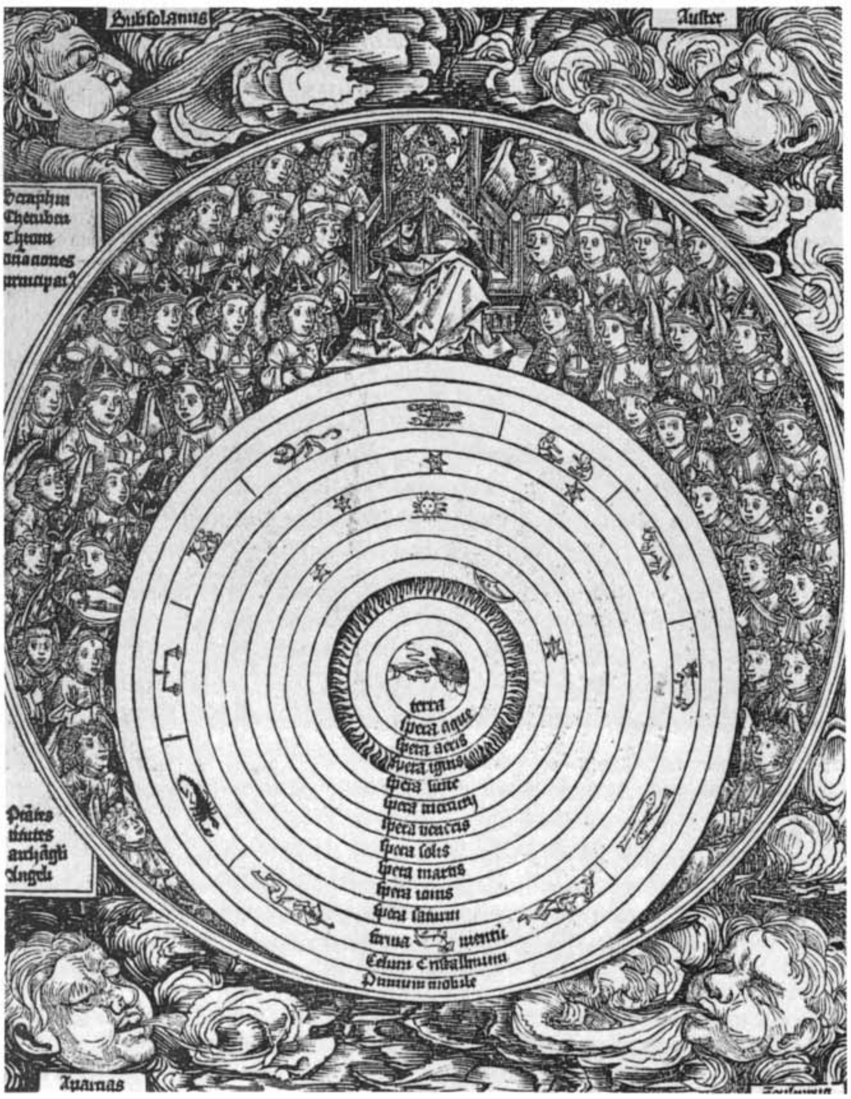
There Columbus faced a panel of clerics, "an imposing array of professors, friars and dignitaries of the church" who "came prepossessed against him, as men in place and dignity are apt to be against poor applicants." They ridiculed the idea of the roundness of the earth and quoted Scripture to infer its flatness. Columbus, a profoundly religious man, found himself in danger of being convicted not only of error but also of heresy.

In reality, knowledge of the earth's round shape was always part of the Western heritage. As the Dark Ages had waned, and with the rediscovery of the works of Aristotle, the notion of a spherical earth entered the curriculum of the newly founded universities. Sacrobosco's *Sphaera*, written in the 13th century and today holding the record as the astronomy textbook with the most editions, gave a simple argument for the spherical shape of the earth in a north-south direction: travelers found that the Big Dipper and Pole Star rose higher in the heavens as they traveled north. As for the east-west direction, Sacrobosco gave a different, more subtle argument: an eclipse of the moon would take place at a particular, unique moment, regardless of where an observer was located, but from different longitudes the height of the moon in the sky would vary in a fashion compatible with a spherical geography.

The problem Columbus faced in Salamanca, then, was not convincing Isabella and Ferdinand that the earth was round but rather that its size and the extent of the Eurasian landmass made the bold notion of a westward voyage to Cathay and the Indies not too unreasonable. The diameter of the earth had been deduced with fair accuracy by Eratosthenes in ancient Alexandria; his round number of 252,000 stadia converts to a circumference of just under 40,000 kilometers, provided that a stade equals 157.5 meters, as the distinguished historian of astronomy J.L.E. Dreyer has argued. Islamic geodesists had reworked Eratosthenes' measurements; al-Farghani, working in the Baghdad group under Calif al-Ma'mun in the early ninth century, had got the equivalent of 20,400 Arabic miles (40,253 kilometers, as compared with the modern figure of 40,075). Columbus assumed, incorrectly, that the Arabic miles were equivalent to Roman ones, which gave him a circumference of 30,044 kilometers, only three quarters of the actual distance.

In addition, Columbus significantly exaggerated the terrestrial longitude of China and hence its distance from Europe. He reckoned the eastward distance to Japan to be as great as 283 degrees, putting the westward distance from the Canaries under 5,000 kilometers. These two erroneous estimates suited Columbus just fine because they made his daring goal of sailing westward to the Indies seem reasonable.

When the court convened in Salamanca around Christmastide in 1486, the scholars there objected to Columbus's diminished estimate of the size of the earth. The circumference they defended was close to the one we accept today. Without his fictitious estimate, Columbus could not have justified his audacious expedition. The myth of the learned flat-earthers is "pure moonshine"; as the eminent biographer Samuel Eliot Morison remarked: "Washington Irving, scenting his opportunity for a picturesque and moving scene, took a fictitious account of this nonexistent university council published 130 years after the event, elaborated it, and let his imagination go completely." Irving's account is gripping drama, "for we all love to hear of professors and experts being confounded by simple common sense. Yet the whole story is misleading and mischievous nonsense."



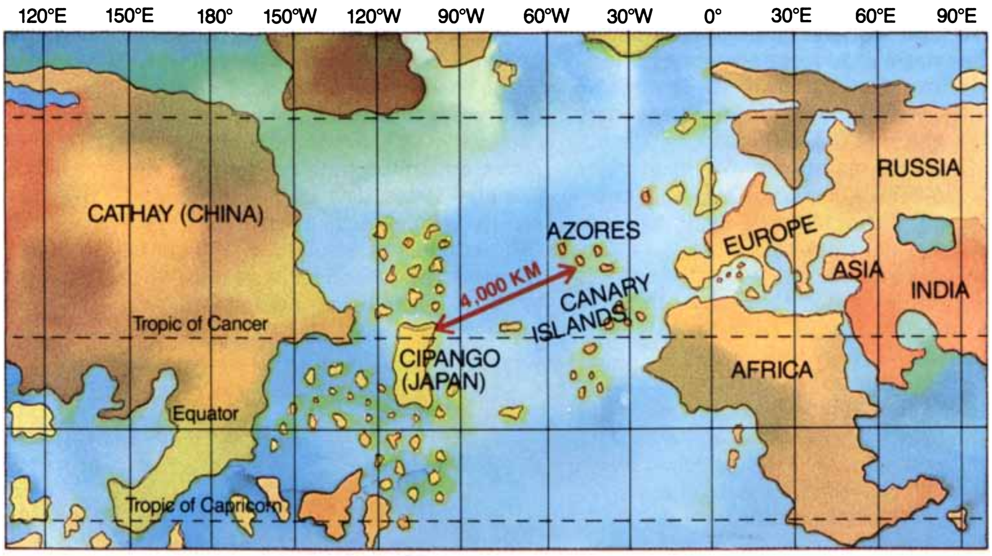
GEOCENTRIC UNIVERSE is shown in this print from the *Nuremberg Chronicle,* a world history published in 1493. The round earth is fixed at the center, surrounded by the planets, including the moon and sun, and by the sphere of fixed stars. In the outermost circle are depicted God and the ranks of angels and the elect. The four winds decorate the corners of the page.

Indeed, except for his wildly mistaken geodesy, Christopher Columbus actually had relatively little to do with astronomy. He is sometimes depicted with stars and primitive navigational devices such as the nocturnal—an instrument for finding the time of night from the position of the Big Dipper—but what little evidence exists concerning Columbus's use of the stars for navigation suggests that he might have got equally good answers just by guessing. Confused by the tropical skies

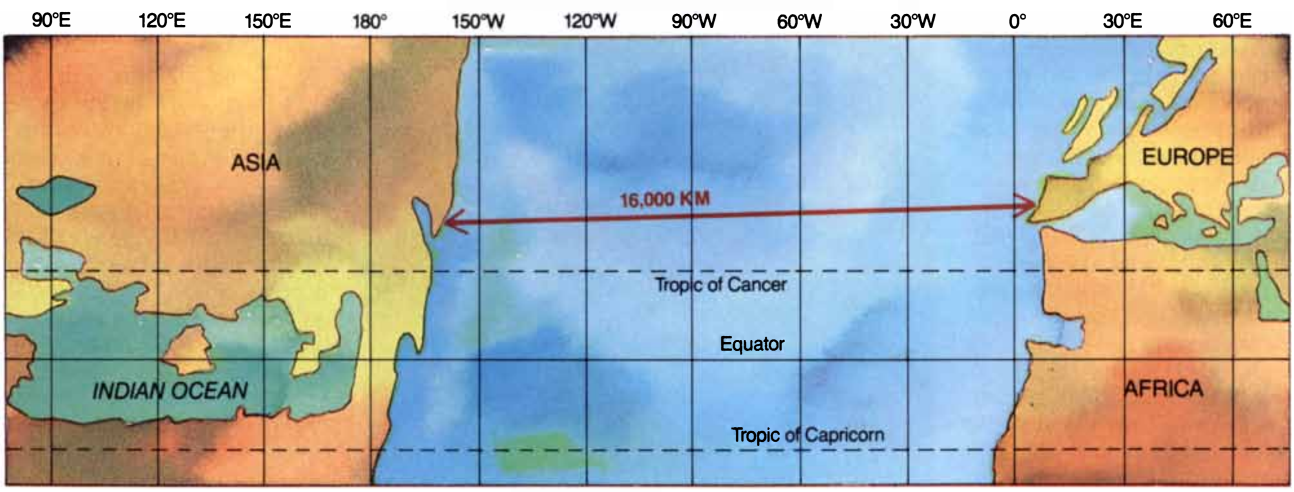
© 1992 SCIENTIFIC AMERICAN, INC

The Earth's Circumference and Continents

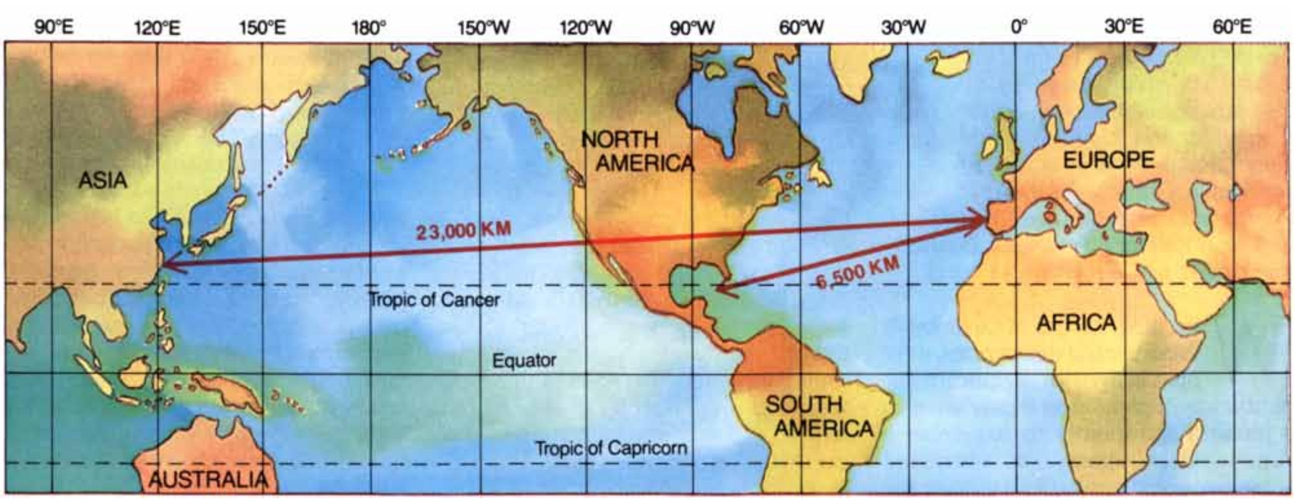
Conflict between Columbus's beliefs about the circumference of the earth and the arrangement of its continents (*top*) and Ptolemaic geography (*middle*) is clear on maps. Only by understating the earth's circumference and overstating the breadth of Asia could Columbus justify a voyage west to the Indies. Both maps of course omitted North and South America (*bottom*). The continents on the top map are taken from a globe made by Martin Behaim in Nuremberg in 1492.



Columbus's View



Ptolemaic View



Modern Map

© 1992 SCIENTIFIC AMERICAN, INC

and the absence of familiar circumpolar constellations, he twice mistook the star Beta Cephei for Polaris, getting a latitude 21 degrees too far north. As Admiral Morison emphasized, Columbus was a dead reckoner, not a celestial navigator.

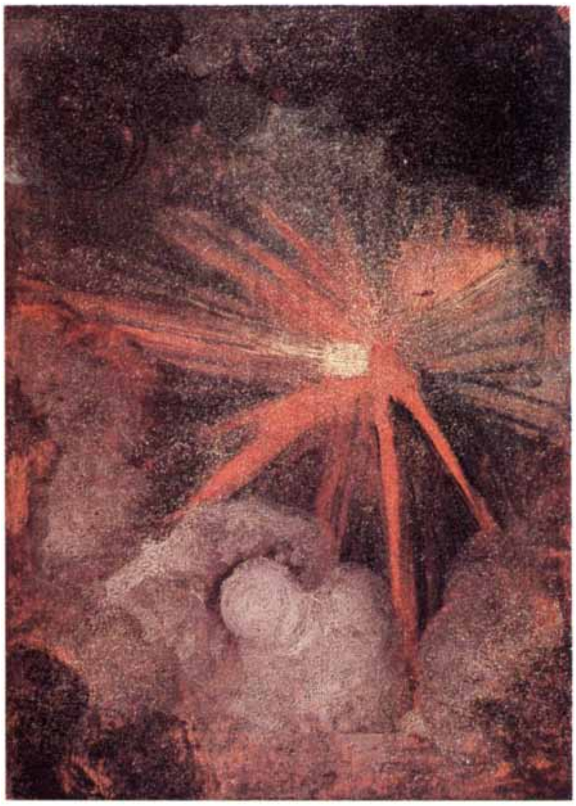
The one conspicuous and legendary exception to Columbus's unastronomical tastes was his use of *Ephemerides,* by Johann Müller (known as Regiomontanus), to foretell an impending lunar eclipse. On his fourth voyage Columbus became stranded on Jamaica, his ships so wormed they were no longer seaworthy. A small party headed east in an open canoe to seek relief from Hispaniola and its capital, Santo Domingo (in the present-day Dominican Republic). The governor of Hispaniola, however, was by no means pleased with the idea of rescuing Columbus. He feared he might be replaced in his lucrative assignment, and so he dragged his feet in sending aid.

Months wore on, and about half of Columbus's men mutinied and tried to sail by canoe to Hispaniola. The Jamaican Indians, now sated with glass beads and other trading trinkets, became increasingly reluctant to provide food for the diminished but nonetheless hungry crew.

From the *Ephemerides* Columbus learned that the moon would be eclipsed on the leap night of February 29, 1504, and he made sure that the Indians knew that the moon would rise dark and bloody as a sign that God was displeased with them. The navigator kept out of sight until the eclipse was over and then came out of his cabin to announce that God had answered his prayers on their behalf. The event so impressed the Jamaicans that they gave Columbus and his crew more than enough food to stave off starvation. (An echo of the episode has been enshrined in American literature in Mark Twain's *A Connecticut Yankee in King Arthur's Court.*)

Why, then, in light of Columbus's astronomical deficiencies, were his voyages so significant for astronomy? Splendid as it may seem to a modern astronomer to remember a time when astronomy was a required topic for every university student, the actual level of Sacrobosco's text was exceedingly elementary. The *Sphaera* described the daily and seasonal motions of the celestial sphere, but it said almost nothing about the motions of the planets.

Knowledgeable medieval astronomers, however, believed that embedded within the Aristotelian spheres was a complex series of subsidiary epicycles and equants, which produced the varying direct and retrograde motions of the planets. These devices had been described by the Alexandrian astronomer Claudius Ptolemy around A.D. 150 in his *Almagest,* a work so technical that virtually no one in Latin Christendom had mastered it. Ptolemy was also the preeminent geographer of his age, and his maps were accepted without question.



***FIREBALL,* painted by Albrecht Dürer in 1492, depicts a stony meteorite that landed near Ensisheim in Alsace. Although the meteorite was arguably the most spectacular astronomical event of that year, the author contends that the discovery of the New World had a far greater effect on astronomical thought.**

During the 15th century, the *Almagest* was at last rediscovered: for the first time there appeared in Europe two astronomers competent enough to understand that fundamental treatise and to criticize its earlier commentators. Regiomontanus and Georg Peuerbach embarked together on an abridged translation of Ptolemy's masterpiece. After Peuerbach's death in 1461, Regiomontanus carried on the project and, in order to secure its widespread distribution, became the first scientific printer. Ironically, Regiomontanus, the greatest mathematician and astronomer of his century, died in 1474, before he could publish either his continuation of Peuerbach's work or his own equally impressive treatise on trigonometry.

Peuerbach and Regiomontanus grasped the details of Ptolemaic astronomy, but they were not pleased with what they saw. In 1464 Regiomontanus wrote a brief but penetrating critique of the theory to a fellow mathematical astronomer. The tables did not give accurate predictions, he reported. He had seen Venus three quarters of a degree out and Mars off by a full two degrees, and a lunar eclipse in 1461 had ended an hour before the calculations indicated. Furthermore, the moon's apparent diameter, according to Ptolemy's theory, should sometimes be twice as large as it is at other times, a phenomenon never observed.

Nicolaus Copernicus, who was born two years before Regiomontanus's death, was also aware of the defects in geocentric predictions of planetary positions. At one point in his notebook he recorded that Mars was two degrees ahead of the tables and Saturn a degree and a half behind [see "Copernicus and Tycho," by Owen Gingerich; SCIENTIFIC AMERICAN, December 1973]. Yet he never mentioned this fault in print, and his own heliocentric-based tables did not correct the errors very effectively.

That Copernicus was seemingly unconcerned by these deficiencies is a very interesting and important point. Despite popular literature to the contrary, errors in the tables of planetary positions had virtually nothing to do with the choice between a geocentric and a heliocentric viewpoint. These two cosmologies were in effect geometric transformations that produced virtually identical predictions: merely transforming to a sun-centered system was insufficient by itself to produce better tables. By the same token, errors in prediction could at least initially be corrected within a geocentric framework as easily as in a sun-centered one.

In fact, Copernicus had no observational proof at all for his new blueprint.

© 1992 SCIENTIFIC AMERICAN, INC

As Galileo was to say a century later, "I cannot admire enough those who accepted the heliocentric doctrine despite the evidence of their senses." Instead the Polish astronomer was guided by an aesthetic vision—a "theory pleasing to the mind."

The heliocentric viewpoint explained why Mars, Jupiter and Saturn appeared to reverse their direction of travel through the sky only when they were in opposition to the sun. In the Ptolemaic system this retrograde motion was an accident of nature, a "fact-in-itself." Copernicus made it a "reasoned fact," and the lack of an explanation in Ptolemy's system became an anomaly. Once the linkages were made, it became obvious why the retrogression for Jupiter was smaller than for Mars and why the retrogression for Saturn was smaller than for Jupiter.

Finally, Copernicus made sense of the mysterious slow displacement of the eighth sphere, the so-called precession of the equinoxes. The discovery of this motion had troubled classical cosmologists. If the earth was suspended in space, however, revolving about the sun, and spinning on its axis, it was not difficult to envision a third motion, a slow, conical displacement of that axis.

These radical innovations laid the foundations on which Galileo, Kepler and Newton built a new model of the heavens. Yet Peuerbach could have made the same geometric transformation a century earlier; the Islamic cosmologists could have made it in the ninth century. Why did the new astronomy wait until the 16th century and the opening decades of the Age of Exploration?

Copernicus lived in an era of rapid change. Perhaps the most visible of those changes was Gutenberg's invention of printing from movable type. With only one known exception, all of Copernicus's documentary sources were printed books. And once his heliocentric cosmology was written down, it was printed in an edition of perhaps 400 copies, guaranteeing wide distribution and ongoing discussion of his ideas.

Then there was the Reformation. Copernicus was a canon in a Catholic cathedral, whereas the young pupil who persuaded him to allow his *De revolutionibus orbium coelestium* to be printed was a Protestant from Wittenberg, the hub of Lutheran activity. It was a time of religious upheaval, when many traditional ideas were under challenge.

But even more to the point, Copernicus lived in an age when courageous seamen were rewriting the time-honored geography of Ptolemy. Copernicus was a student at Cracow when Columbus made his first voyage. News of these discoveries came quickly to Cracow, and to this day the Jagiellonian University preserves the oldest known globe showing the New World. Even if Copernicus had left before the news arrived, he surely heard it soon thereafter while he was pursuing graduate studies in Italy.

The Alexandrian astronomer was probably even better known for his geography than for the geocentric cosmology that bears his name. His *Geography,* written in the second century A.D., left a crucial legacy to cartographers in its instructions for map projections. Reworking information sorted out from travelers' reports and from his predecessor, Marinus of Tyre, Ptolemy had assembled his best estimates of the latitudes and longitudes of locations in the then-known world. These, in turn, became the basis of spectacular atlases eventually published in the 1480s. But by the early 1500s his reputation was fast eroding. Although Columbus had believed himself to be following the old geography, his landing in the "Indies" challenged accepted maps. And when it became clear that he had indeed discovered a new continent, the classical globe was clearly obsolete. If Ptolemy's geography had fallen by the wayside, could not his cosmology also be questioned?

The defects Regiomontanus saw in classical astronomy for the most part went uncorrected by Copernicus. Yet the heliocentric blueprint was the single most essential step for the ultimate reform of astronomy. It offered a wrenching realignment of human thought, and it paved the way for the brilliant technical achievements of Kepler and Galileo.

Copernicus's *De revolutionibus* was published in Nuremberg in 1543, in a world already prepared for change. In 1566 the Basel publisher Henricpetri issued a second edition of the work. Among those who obtained the reprint was Thomas Digges, who became the first English astronomer to convert to the new cosmology. Above the title on his copy he wrote, "*Vulgi opinio error,*" "the common opinion is wrong," meaning that he no longer accepted the time-honored notion that the earth was fixed at the center of the universe.

In offering an English translation of its key cosmological passages, Digges wrote, "I thought it convenient to publish this, to the ende such noble English minds (as delight to reache above the baser sort of men) might not be altogether defrauded of so noble a part of Philosophy." He appended a magnificent heliocentric diagram to his presentation, which contained a novel feature: no longer were the stars fixed to a distant shell, but they were spread out toward infinity. "And therefore," he concluded, "immovable." Presented in 1576, this model was a mind-boggling conception, an astonishing step from the closed world of the ancients toward today's vast universe.

Digges and several of his contemporaries, including Kepler's teacher, Michael Maestlin, searched hard for empirical ways to confirm the sun-centered planetary arrangement, but in vain. It remained a leap of faith but a compelling aesthetic vision to those who understood its unity. The faithful also had to discard the long-accepted Aristotelian physics, which predicted that birds and clouds would be left far behind as the earth spun on its axis. As another contemporary astronomer, Tycho Brahe, declared, "Copernicus nowhere offends the principles of mathematics, but he throws the earth, this lazy, sluggish body, unfit for movement, into a motion as swift as the aethereal torches [the stars]."

In the absence of observational proof, the adoption of Copernicus's blueprint required a climate of opinion willing to accept new ideas and no longer locked into hoary traditions in which ancient learning stood on a pedestal. Columbus helped to provide that new intellectual climate. His empirical evidence decisively demonstrated the incompleteness of Ptolemy's geography and so prepared the way for a revised understanding of the place of the earth in the cosmos. The old views were crumbling. By 1611 John Donne would write, "And new Philosophy calls all in doubt,/ The Element of fire is quite put out;/ The Sunne is lost, and th'earth, and no man's wit/Can well direct him where to look for it."

FURTHER READING

ADMIRAL OF THE OCEAN SEA: A LIFE OF CHRISTOPHER COLUMBUS. Samuel Eliot Morison. Little, Brown, 1942.

INVENTING THE FLAT EARTH: COLUMBUS AND MODERN HISTORIANS. Jeffrey Burton Russell. Praeger, 1991.

WHEN DO ANOMALIES BEGIN? Alan Lightman and Owen Gingerich in *Science*, Vol. 255, pages 690–695; February 7, 1992.

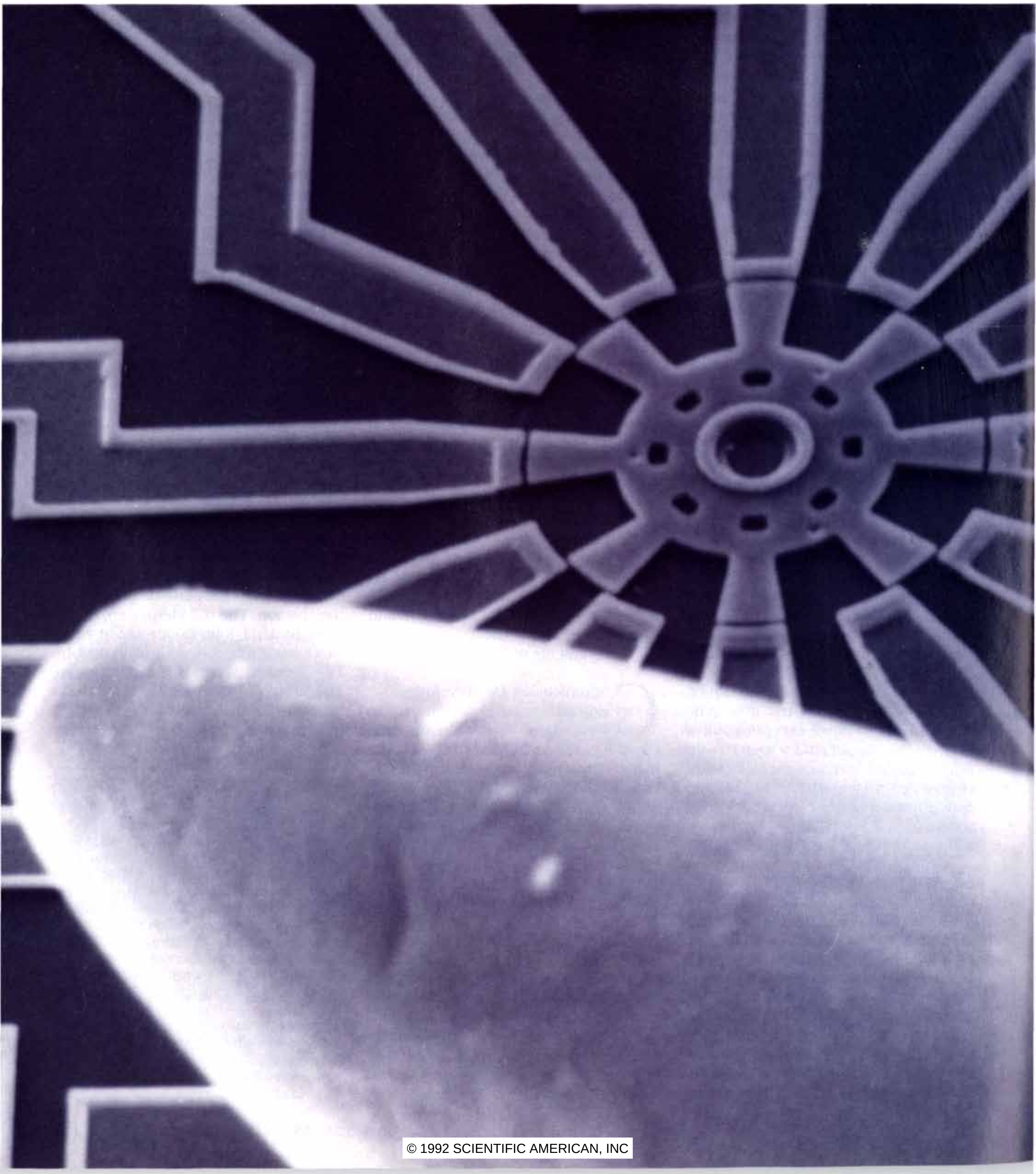
THE METEORITE OF ENSISHEIM: 1492 TO 1992. Ursula B. Marvin in *Meteoritics*, Vol. 27, No. 1, pages 28–72; March 1992.

COLUMBUS AND AN ECLIPSE OF THE MOON. Donald W. Olson in *Sky and Telescope*, Vol. 84, No. 4, pages 437–440; October 1992.

© 1992 SCIENTIFIC AMERICAN, INC

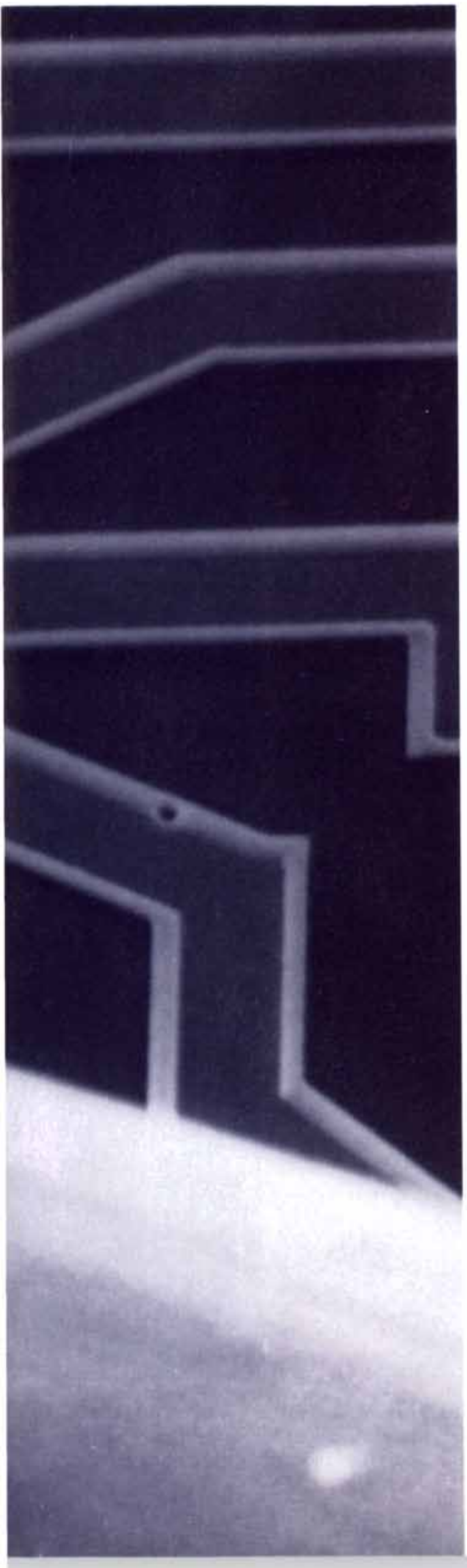
MICRON MACHINATIONS

by Gary Stix, *staff writer*



© 1992 SCIENTIFIC AMERICAN, INC

Silicon is becoming both bricks and mortar for armies of gears, valves, pumps and sensing devices that may turn the surface of microchips into diminutive factories and laboratories.



The Central Intelligence Agency was out of luck: someone had forgotten the microscope. A group of technical types from the agency was huddled around the microchip that Case Western Reserve University professor Mehran Mehregany was holding in his hand. But Mehregany's little eyepiece magnifier could not resolve what was going on down there on the surface of the chip.

The CIA had not really told Mehregany why it had asked him to visit Washington to describe his work. Even so, Mehregany wanted to prove to his audience that his research was more than just talk, that it is possible to make machines so small that more than 1,000 of them could fit inside the block letters "CIA." Without the requisite microscope, the little dark dot on the smooth silicon surface could just as well have been a speck of dust as a spinning motor.

Had the CIA staffers been able to see just a bit more clearly, they would have made out a rotor element that looked a little like a miniaturized replica of wheels that could have driven a 19th-century mill. And if Mehregany and a few hundred other researchers are right, microscopic machines may presage a new industrial revolution—one that could eventually elevate a gear or a pump the size of a healthy protozoan to a place alongside the most celebrated valve of all time, the transistor.

The nascent field of micromechanics envisions "smart" pills that could inject dosages of drugs inside a patient with split-second precision. An array of positioning arms, each element having a cross section that spans less than a micron, might maneuver across a square inch of disk storage space, reading and writing enough data to accommodate every edition of the *Encyclopaedia Britannica* ever printed with lots of room to spare. A 40-pound mass spectrometer, which can be used as a general-purpose gas sensor, could be reduced—vacuum pumps, detectors and all—to the dimensions of a pocket calculator.

The list is beginning to become more than just the imaginings of futurists and technophilic dreamers. The field's international, biannual gathering of the tribes, called Transducers, produced a technical document in 1991 that weighed more than the Manhattan White Pages. The 1,089-page tome constitutes a diverse record: a device moved by bubble pressure, a microscopic tweezers, a probe to detect nerve signals while lodged inside the brain and a sensor that, like the human hand, can distinguish between hard and soft. One version of Mehregany's micromotors, which were also listed in the table of contents, might someday serve as the muscle, or actuator,

EARLY MICROMOTOR, with a 130-micron-diameter rotor, is nearly hidden by a pinpoint. Mehran Mehregany made the motor while he was still a graduate student at the Massachusetts Institute of Technology. He continues to pursue the work as a professor at Case Western Reserve University.

© 1992 SCIENTIFIC AMERICAN, INC

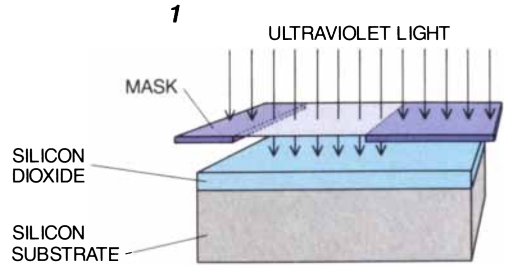
for some combination of sensors, valves and electronics on a single silicon chip—capital equipment for a microchemical plant or a cluster of scientific instruments at a scale of a few microns.

At present, these vanishing engines are rather frail devices. In seminars such as that held for the CIA, Mehregany must constantly field questions about the usefulness of a motor that produces substantially more torque than the flagella of a salmonella bacterium but still a tiny fraction of that supplied by a watch mechanism. And one that required Mehregany's graduate students, Vijay Dhuler and Keren Deng, to stay up for two nights to bond the wires from the little machine to a comparatively enormous battery and its control circuitry.

A solution may lie beneath the microscope's lens. With magnification, the CIA officials would have seen a few wires connected to just one motor. Beside the working machine were 99 others that remained stationary because their electrodes lacked the requisite connections. Mehregany calculates that up to a million motors could be produced on a large silicon wafer from which the individual chips are cut. An ensemble of a few hundred thousand motors could pump nearly a liter of fluid each minute. That is a potentially useful amount, although Mehregany still muses on what to do with a vast garden of micropumps.

Micromotors will also require more than an overnight jump-start by two red-eyed graduate students. For that, microelectronics may help. Photolithography combined with etching by a chemical plasma has supplanted stamping and casting as the predominant mass-production technology of the late 20th century. Methods for placing thousands of wires and transistors in a space dwarfed by Lincoln's right nostril on a postage stamp can also fabricate submillimeter spinning and vibrating structures, machines that even Henry Ford would have recognized.

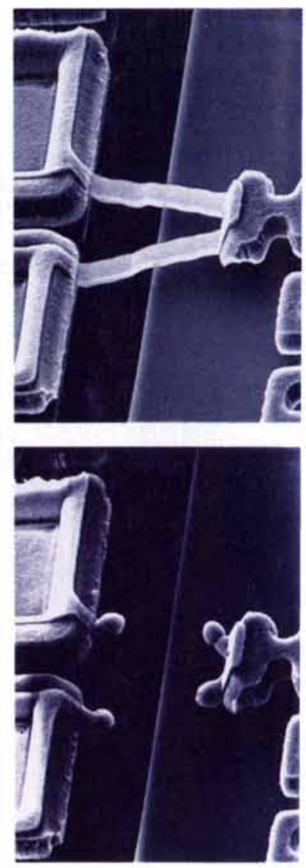
To some, silicon seems an unlikely structural material, the main constituent of sand castles, not tiny motors nor little cantilevers whose vibrations measure acceleration. In 1981 Kurt E. Petersen, then an IBM researcher and now the vice president of technology at Lucas NovaSensor in Fremont, Calif., wrote a seminal technical paper, "Silicon as a Mechanical Material." The hardness of



BEAM BUILDING ON A CHIP was a graduate project for Roger T. Howe. Ultraviolet light streamed through a photolitho-

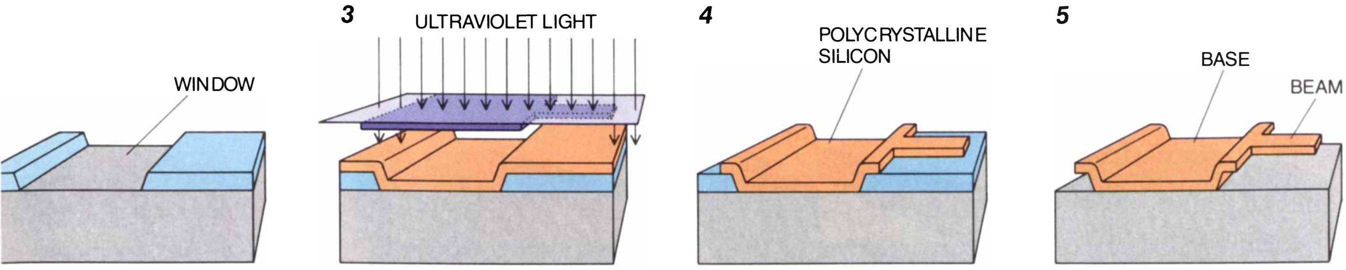
silicon, he noted, "is close to quartz, just below chromium, and almost twice as high as nickel, iron, and most common glasses." In addition, silicon's tensile strength exceeds that of steel, although its brittleness keeps it from being a wonder material in all respects.

The micromachines being constructed by Mehregany and others are an outgrowth of more than 20 years of research, development and commercialization of sensors made mainly from silicon. In the oldest process, areas of single-crystal silicon that have first been



ROGER T. HOWE poses with graduate students Clark Nguyen (*center*) and Weijie Yun (*right*) at the University of California at Berkeley. Howe pioneered the surface micromachining process and has gone on to develop methods for assembling these structures: a small current courses through a fuse two tenths of a micron thick (*center top*) that releases a folded beam (*cen-*

© 1992 SCIENTIFIC AMERICAN, INC



graphic mask (*1*) before hydrofluoric acid etched a "window" in a layer of silicon dioxide (*2*). A layer of polycrystalline silicon was deposited from a vapor before patterning the base and beam structure (*3*). A chemical plasma etched areas of lithographically exposed polycrystalline silicon (*4*). Acid then removed, or "sacrificed," remaining silicon dioxide (*5*).

exposed through a photolithographic mask are eaten away by alkaline chemicals. Etching produces concave, pyramidal or other faceted holes, depending on which face of the crystal is exposed to the chemicals. These sculpted-out cavities can then become the building blocks for cantilevers, diaphragms or other structural elements needed to make devices such as pressure or acceleration sensors.

This technique, which first emerged in the 1960s, has come to be known as bulk micromachining because the chemicals that pit deeply into the silicon produce structures that use the entire mass of the chip. At about the time Petersen wrote his article, a graduate student at the University of California at Berkeley was butting up against some of the limitations of bulk micromachining—a problem that led to a method of making sensors and actuators that more closely parallels the conventional process for making integrated circuits.

For his master's degree project, Roger T. Howe, under the tutelage of his professor, Richard S. Muller, was planning to make a gas sensor—a small vibrating beam whose frequency would shift as a gas vapor condensed on its surface. Howe wanted to make the device with bulk micromachining but became discouraged after discussing his plans with Petersen and other researchers. Placing capacitors underneath the tiny beam to make it vibrate and to sense the resonance turned out to be an exceedingly difficult task.

In pondering the problem, Howe remembered that polycrystalline silicon (in which faces of the crystal line up randomly) is routinely deposited onto a silicon substrate from a gas vapor to form a transistor element, a gate, that switches the device off and on when a voltage is applied. So he turned to "poly," as it is called, to build his tiny cantilever.

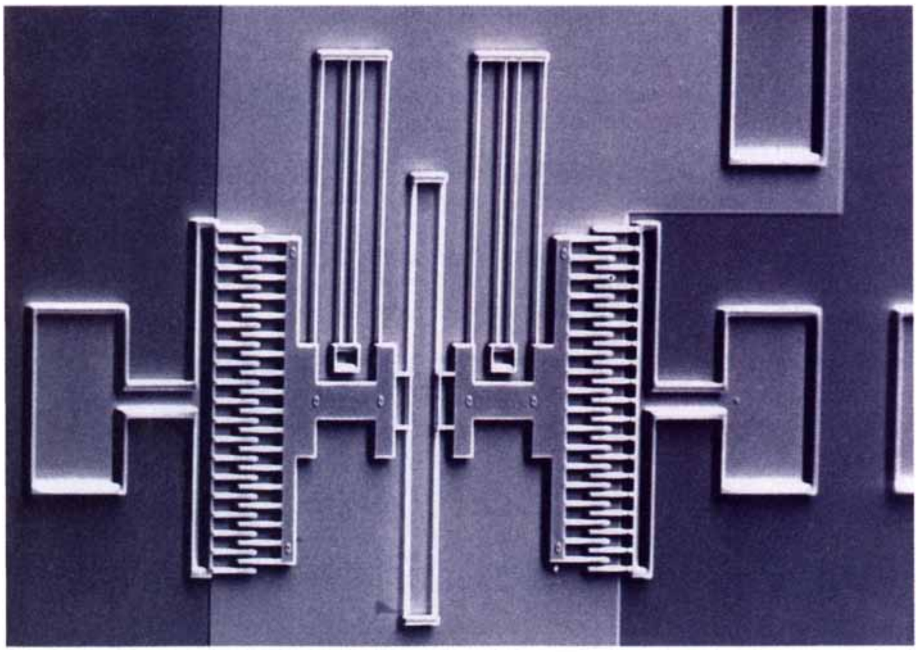
Diving Board on a Chip

Howe deposited polycrystalline silicon from the gas vapor onto a silicon substrate. It was patterned with photolithography and then etched before hydrofluoric acid removed, or "sacrificed," a layer of silicon dioxide to leave the suspended beam. Howe went on to complete what had become a doctoral project by making a gas sensor.

What was most significant about Howe's tiny diving board was not the gas sensor but how its fabrication relied on the vapor deposition and etching technology used to manufacture the most ubiquitous type of transistor, the metal-oxide semiconductor. The alkaline chemicals for bulk micromachining, foreign to conventional chip processing, were nowhere to be seen.

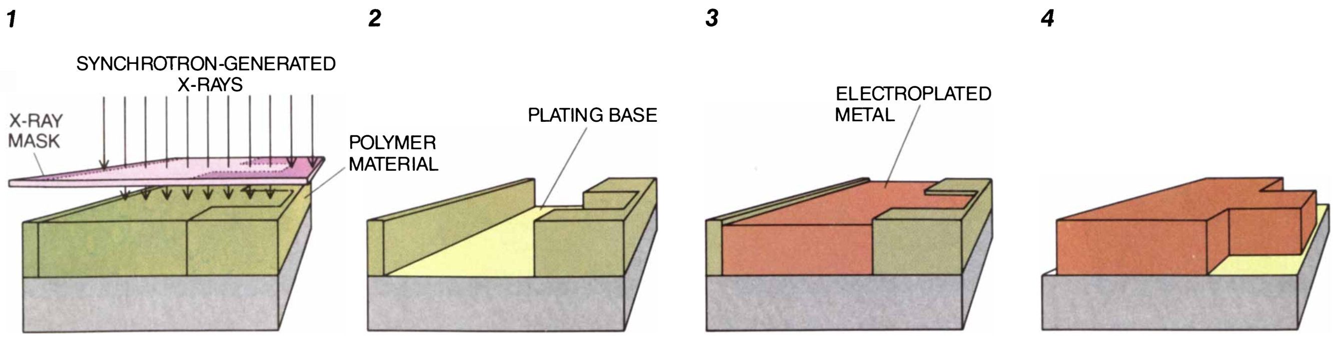
The technique is called surface micromachining because it deposits a film of silicon a few microns thick, from which beams and other edifices can be built. The thinness of these structures is a challenge to the designer, who must derive useful work from machines whose form is essentially two-dimensional. Since Howe completed his gas sensor, others have used the technique to build valves, motors and a strange armamentarium of devices that bear more connection to a medical textbook than to the mechanical engineering curriculum. The gray and black imagery of the scanning electron microscope reveals inchwormlike chains of parallelograms that expand and contract, arrays of cilialike elements that curl up from the plane and push-pull contraptions whose design was influenced by muscle fiber.

Howe, who is today an intense 35-year-old professor at Berkeley, has added still other fabrication methods that work the surface of a silicon chip as if it were a piece of sheet metal. At a re-



***ter bottom*). This process could help fabricate a mechanical acoustic filter (*right*) a few hundred microns wide. The filter couples vibrations from one set of folded beams to another through a spring, the rectangular structure in the middle.**

© 1992 SCIENTIFIC AMERICAN, INC



INJECTION MOLDING of microscopic parts can be carried out with a process developed at the Karlsruhe Nuclear Research Center in Germany. Called LIGA, it stands for the German acronym for *Lithographie* (lithography), *Galvanoformung* (electroplating), *Abformung* (molding). X-rays from a synchrotron penetrate the transparent part of a lithographic mask (*1*). The exposed part of the polymer layer, up to several hundred microns in thickness, is removed using a developing chemical

cent technical conference, he and his students presented papers on automated assembly techniques for micromechanical structures. Hundreds of tiny silicon fuses, two tenths of a micron thick, burst in unison when exposed to a current, releasing a suspended beam. Aluminum weld joints over the entire chip melt when a current is applied, fastening down the beam.

These processes might eventually help put together Howe's mechanical structures, one of which is a miniature of a decades-old electrical device. Originally, resonating inch-size plates made of nickel-iron alloys served to filter acoustic signals, only to be supplanted by capacitor elements in integrated circuits that were cheaper and smaller, although they did not narrow a signal as efficiently. Now Howe and his graduate students have come up with micromechanical filters that can produce more signal and less noise and are able to tune a wider range of frequencies than can the integrated elements that they replace. Hundreds can take up residence on a centimeter-square plot that a microchip occupies.

The filter works by using the voltage from an audio signal, and perhaps one day video, to move a series of suspended bars back and forth between the stationary plates of an electrode, a structure that is often compared with two intertwined combs. The oscillations, transferred by a microscopic spring, alter the vibrations of an adjacent set of comblike digits and electrodes. A filtered signal results from coupling the resonances between the two "comb drives," which then get converted back into an electrical signal. "The integrated circuit is becoming inefficient compared with a few coupled resonators," Howe asserts.

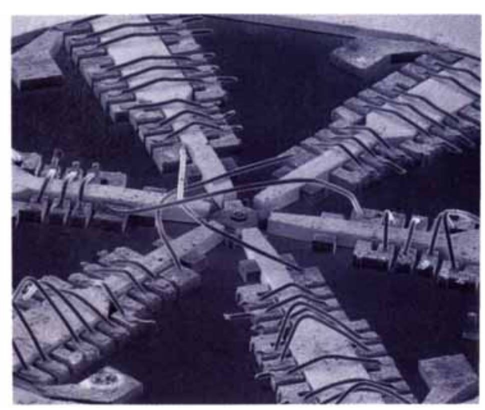
Howe's creativity also helped to conceive of a more well known but perhaps less functional product of micromechanics research, one of the first rotating micromotors. Howe was among those who came up with the idea for one version of the micromotor while teaching at the Massachusetts Institute of Technology in the mid-1980s. The effort to create a micromotor quickly became a heated race between M.I.T. and Berkeley. (Howe's return to Berkeley before either side declared itself a winner created a minor furor, although he says he returned to California to be near his family and never worked on the micromotor project there.) Berkeley ultimately won the race, proclaiming, in the summer of 1988, that it had successfully powered a micromotor with a series of capacitors that generated an electric field to move a rotor.

The micromotor immediately became an icon for the young field. Micrographs of beams, cantilevers and comb drives look appropriately like something that one expects to see under a microscope. But when it was pasted across newspapers worldwide, the spinning rotor of the micromotor had a strange but familiar look. It was a shrunken version of an archaic-looking machine, something that might have tormented Charlie Chaplin in *Modern Times*.

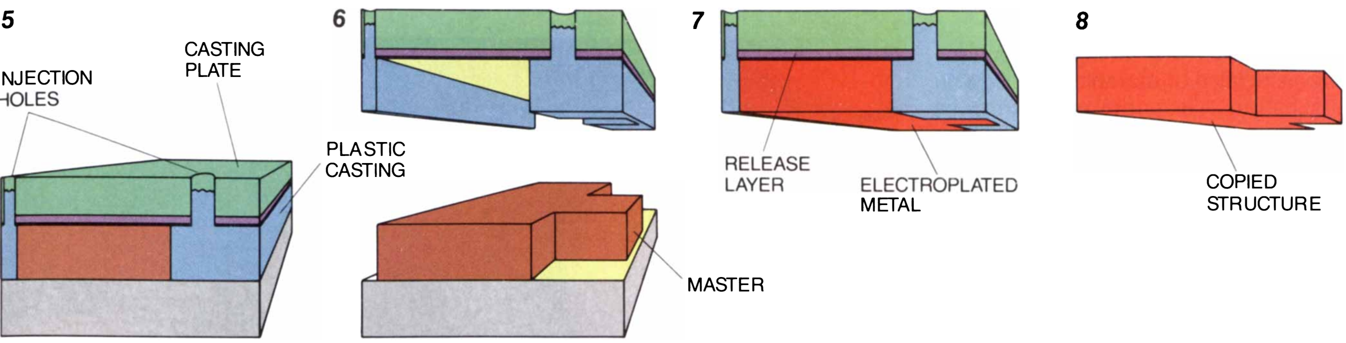
The first micromotors were exceedingly fragile creatures, ideally suited, in fact, to be the props in a Chaplin movie about technology run amok. Initially, getting them to spin at all was a problem. The rotors, once released by hydrofluoric acid from the silicon substrate below, would often remain rooted by friction to the silicon surface. The motors also had an annoying tendency to run for a few thousand cycles and then abruptly lock up. And residual stresses between the layers deposited in surface micromachining would make the rotors curl up, the "potato chip" effect.

Those early motors can prove an embarrassment to entrepreneurs who have to explain to a customer or financial backer how their microvalve or sensor differs from what some perceive to be a toy. In one sense, though, the micromotors are a success. Making a rotor turn on a bearing is probably the most difficult research task for a would-be micromechanic. A scientist at Bell Laboratories, which also eventually demonstrated its own micromotor, once told a group of researchers there that the voltages required for generating the electric field to turn the rotors would be so high as to render the devices inoperative. It was also thought that the micromotors would become bantam air cleaners that would be buried under a swarm of dust particles attracted by the electric field.

Neither of these predictions turned out to be true, and early problems that seemed to confirm doomsayers' prophecies about the incapacitating influences of friction were also overstated. Certain-



© 1992 SCIENTIFIC AMERICAN, INC



(*2*). Nickel or another metal is electroplated onto a base (*3*). After the remaining polymer is removed (*4*), a casting plate is placed over the metal, and plastic is injected through the plate's holes (*5*). The metal master is then removed and is used to make other molds (*6*). The plastic casting is electroplated (*7*), and the plastic and a release layer are etched away to free the metal structure (*8*). Because LIGA uses lithography, thousands of masters and molds can be made on a substrate.

ly, a theoretical understanding of friction on the microscale is lacking. But Mehregany, who worked on the original M.I.T. team as a graduate student, has continued to pursue research on micromotors vigorously at Case Western.

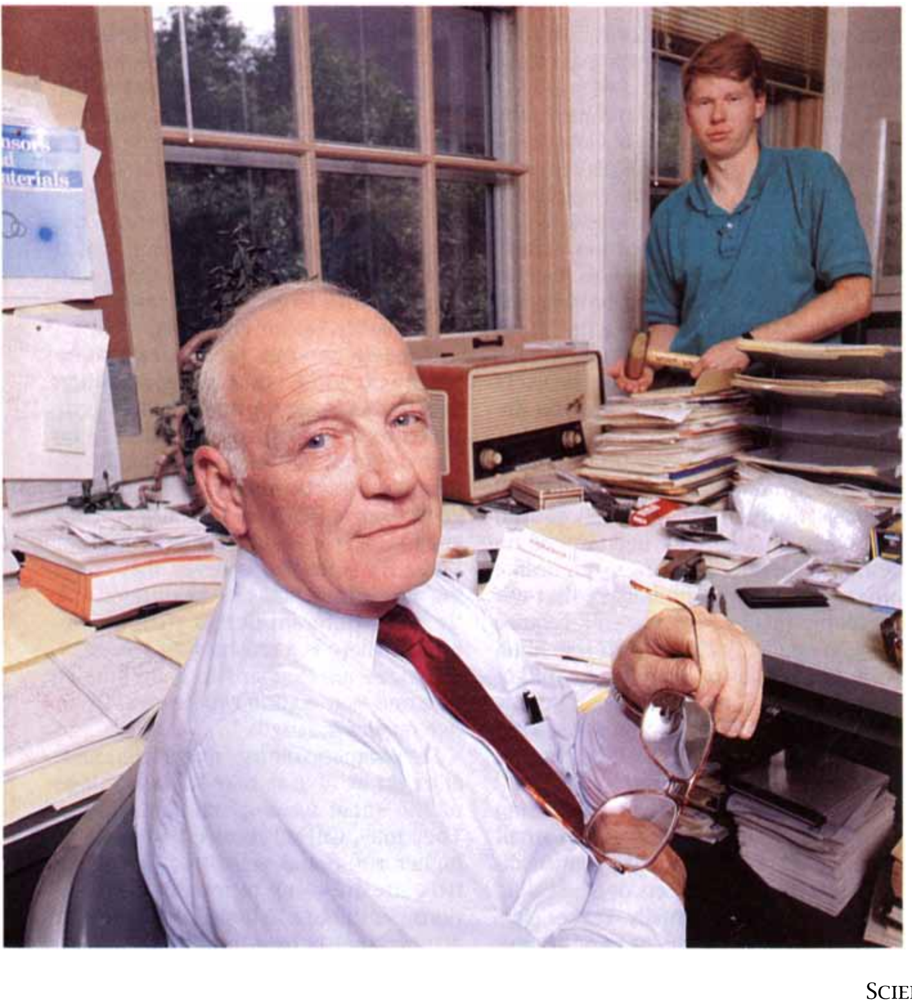
His approach is an empirical one. Mehregany has alternated different designs of bearings and bushings to reduce the amount of friction to about 10 percent of total torque, still from 10 to 100 times more than the percentage found in, say, a washing machine motor. Mehregany's motor has operated at up to 15,000 revolutions per minute. At lower speeds, it worked for days on end until one of his graduate students got tired of baby-sitting the twirling mote. Better materials and control electronics should double the top speed and nudge the frictional component lower. "Friction is no longer a problem," Mehregany proclaims with a touch of overconfidence.

Getting more than a few trillionths of a Newton-meter of torque from a micromotor becomes an easier proposition if the device is not as flat as a pancake. The amount of torque increases with the height of the structure; a surface-micromachined motor is only two or three microns thick. To gain the needed depth, some researchers are turning to other structural materials than silicon.

In the early 1980s Wolfgang Ehrfeld headed a research team at the Karlsruhe Nuclear Research Center in Germany that devised a means to make microstructures that are thicker than they are wide: a nickel device five microns wide and 300 microns tall, for example. Ehrfeld built his work on a process developed at Karlsruhe for making nozzles whose curving shape acts like a minicentrifuge, allowing for separation of uranium isotopes. The technique is known as LIGA, the German acronym for *Lithographie, Galvanoformung, Abformung.*

Like surface and bulk micromachining, LIGA relies on lithographic patterning. But instead of the ultraviolet light streaming through a photolithographic mask, this process utilizes high-energy x-rays that penetrate several hundred microns into a thick layer of polymer. Exposed areas are stripped away with a developing chemical, leaving a template that can be filled with nickel or another material by electrodeposition (*Galvanoformung*). What remains may be either a structural element or the master for a molding process (*Abformung*).

As with surface micromachining, LIGA structures can be processed to etch



HENRY GUCKEL, with mallet-wielding senior graduate student Todd Christenson, has used LIGA to make a rotating magnetic micromotor at the University of Wisconsin–Madison [*see right photograph on opposite page*]. Christenson uses a synchrotron in the wee hours to make gears (*center*) and occasional spoofs, like the 200-micron-long wrench, from whose handle the word "Craftsman" was inadvertently omitted (*left*).

© 1992 SCIENTIFIC AMERICAN, INC

away an underlying sacrificial layer, leaving suspended or movable structures on a substrate. The entire process can be carried out on the surface of a silicon chip, giving LIGA a degree of compatibility with microelectronics. So far Karlsruhe has produced plastic optical devices as well as metal acceleration sensors, turbines and motors—three-dimensional versions of some of the structures fabricated with surface micromachining.

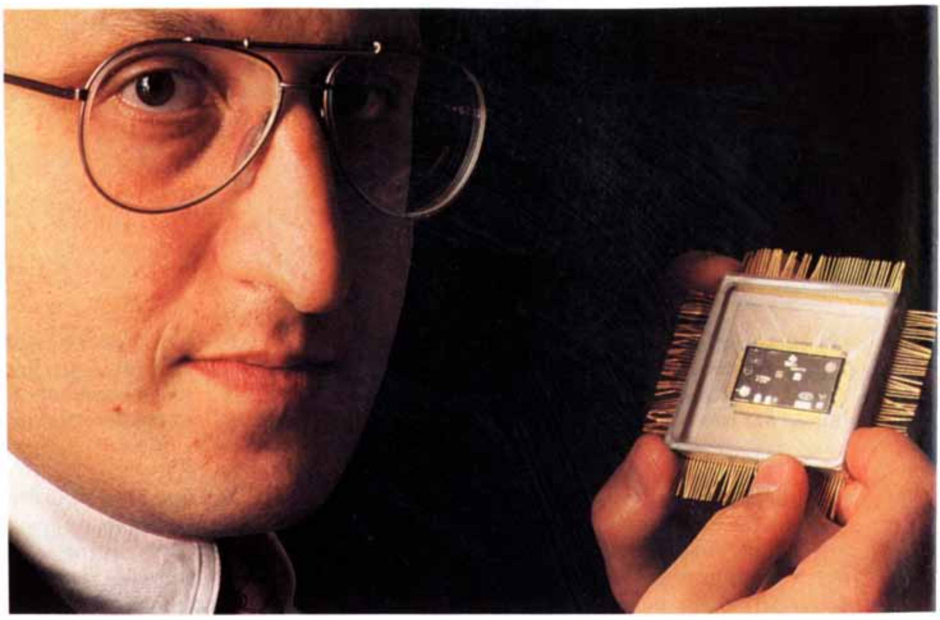
The Big LIGA

LIGA is at the forefront of a government-supported bid to place Germany in the lead in micromechanics, an effort to make up for the country's failure to establish a leading world position in electronics. As research on new nuclear reactor technology has slowed, Karlsruhe plans to deploy more than 200 researchers from the institute for work in LIGA and micromechanics. Meanwhile Ehrfeld, the LIGA pioneer, left Karlsruhe about four years ago when he was not offered a promotion. He is now setting up a competing program at a government-supported institute in Mainz.

Karlsruhe also helped to transfer this technology to private industry. It has licensed its patents to MicroParts, a private concern jointly held by steel, chemical and power-generation companies, which sells LIGA components, the only commercial source anywhere.

Micromechanics in Germany is more than just LIGA. The government also funnels $70 million a year into what it calls microsystems, a hodgepodge of technologies that includes both micromechanics and electronics. This program is administered by VDI/VDE Technology Center, a German institute that is coordinating funding for micromechanics. The Microsystem Technology program is an umbrella for hundreds of small and medium-size companies and research institutes from all over the country. Karlsruhe and VDI/VDE have also spearheaded a proposal for establishing a $1-billion effort within the European Community that would focus on these miniaturization technologies.

A thorn in the side of the Karlsruhe researchers is Henry Guckel, a 60-year-old professor of electrical engineering at the University of Wisconsin–Madison, who claims to have done work that the Karlsruhe team has yet to duplicate. During the early 1980s, Guckel began studying polycrystalline silicon as a material for making suspended or moving structures for sensors. He began to read—or "read between the lines," as he puts it—the papers coming out of Germany on LIGA, all the while noting the synchrotron on his own campus that was used to study x-ray lithography for microcircuits. For the past three years, he has experimented with LIGA.



KAIGHAM J. GABRIEL, head of a micromechanics program at the Defense Advanced Research Projects Agency, holds a silicon chip from Texas Instruments that has two million moving metal mirrors that modulate the brightness of a projection display. As a researcher, Gabriel became a proponent of using arrays of microscopic structures, such as the ciliary motion system devised by Hiroyuki Fujita at the University of Tokyo (*center*). DARPA has funded Cornell University to make an acceleration sensor using opposing scanning tunneling microscope tips. The top tip will be suspended on beams with a 200-nanometer cross section (*right*).

A native of Hamburg who came to the U.S. in 1950 at the age of 18, Guckel will often try to deflate what he perceives as the seriousness of German researchers, attributing Wisconsin's success to the Europeans' extended vacation schedules. "I'm encouraging them to take an extra 10 days of vacation," he chortles. Ever the jokester, he keeps a mallet on his cluttered desk and constantly asks students toiling on graduate projects that exceed half a decade: "Are you finished yet?"

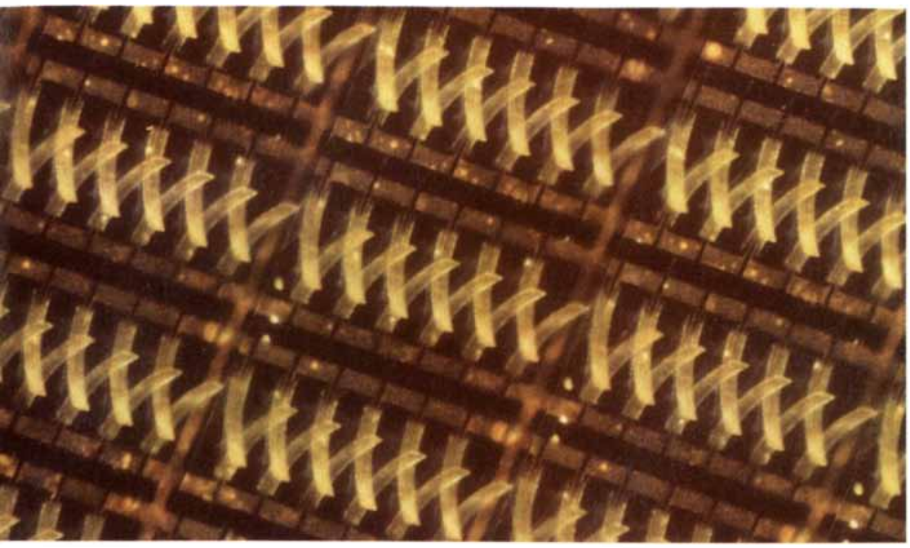
Guckel's group was the first to make sets of toothed nickel gears, 50 to 200 microns in diameter and 200 to 300 microns high, that have been linked together to form gear trains. And, unlike other researchers who believe that microscopic mechanical systems require designs specially adapted to the small scale, Guckel has recently built a micro-Tinkertoy of a motor powered not by static electricity but by the same kind of magnetically driven electric motors that run a kitchen refrigerator.

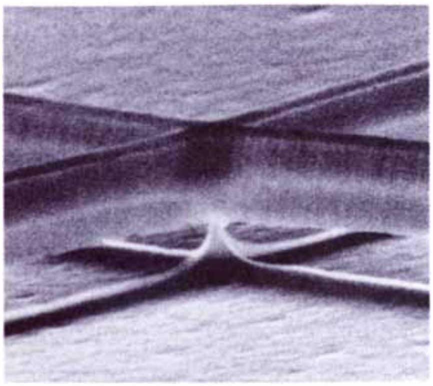
Other researchers have avoided using electromagnetic forces to propel small gears because they thought that at dimensions of 100 microns or so, scaling laws would make normally weaker electrostatic forces equivalent to or greater than magnetic forces. Guckel believes the scaling arguments may be wrong: operating an electrostatic motor at the low voltages used to power a semiconductor chip requires very close tolerances, gaps of less than a micron between the rotor of the electrostatic motor and the electrode that generates the electric field. A magnetic motor, which is driven by a current, not a voltage, can function with larger gaps at reasonable power levels, Guckel contends.

Magnetic motors, though, cannot be used with some types of semiconductor circuits. And, for now, the tiny gear trains turned by the motors have to be painstakingly assembled by hand. "A microsystems technology for four-shift gearboxes is ridiculous," charges Wolfgang Menz, who heads the LIGA effort at Karlsruhe and who believes that copying larger machines is misguided. "A fly has a different design than an elephant. There is a reason that you can't minimize an elephant to a millimeter size and you couldn't increase a fly to five meters in length."

But magnetic motors made from metal might be able to take full advantage of the small mass of micromachines. They may, Guckel thinks, be driven to higher rotational velocities than electrostatic ones—up to a million revolutions per minute. "I hope I don't have to achieve that before I graduate," quips

© 1992 SCIENTIFIC AMERICAN, INC





Todd Christenson, Guckel's senior student. Christenson is working on a multigeared dynamometer that could detect changes of just a few revolutions in gears turning hundreds of thousands of times a minute. The device would provide a means of studying the small-scale effects of friction and other forces.

Guckel's main goal is to move toward a product someone will buy. The U.S. Army has expressed interest in a tiny turbine and generator whose energy would be supplied by waving the device around in the air, picking up enough energy to power a microchip. Still more compelling is the market for disk drives. The IBM Almaden Research Center in San Jose, Calif., has a research group that is investigating small actuators for positioning disk-drive heads and small motors for spinning the disk surfaces. It is exploring a cooperative effort with Wisconsin. Guckel is fascinated by the approximately $50 billion garnered by the disk-drive industry every year. "I tell my students that I expect them to be able to absorb about 10 percent of that for my retirement," he marvels.

Guckel's playful demeanor belies some of the funding problems with LIGA and with other U.S. micromechanics research. High-energy synchrotrons for x-ray lithography are research tools that in the U.S. can be counted on the fingers of one hand. Graduate students at Wisconsin zap their polymer sheets from midnight until shortly after 8 A.M., when they have to yield the beam to another research team.

Guckel points out that LIGA could produce thousands of inexpensive parts through an injection molding process pioneered at Karlsruhe. Even so, the tremendous costs will have to be shared. Guckel has been involved with Ford and other companies interested in establishing a cooperative organization for developing LIGA. But the group is still, as one federal official characterized it, a "debating society."

Wisconsin is now the only U.S. university involved with LIGA. But Louisiana Tech University has just begun to establish a major research program in micromechanics that includes the technique. The $10 million procured for the Institute for Micromanufacturing from the 1992 federal budget by Senator J. Bennett Johnston of Louisiana was intended to help move the state beyond economic dependence on oil and gas. But the level of funding, which is more than the entire field receives annually from the National Science Foundation and other government sources, has bred resentment among more prominent research groups that struggle each year for renewal of $50,000 peer-reviewed grants.

The U.S. arguably retains a leading role in academic research in micromechanics. Universities with programs in the technology have seeded graduates at such corporations as Ford, IBM and Hewlett-Packard. But a broad cross-industry commitment is lacking.

In the meantime, Japanese corporations have begun to emerge as a source of innovation. Their researchers presented 21 technical papers at a major conference last year, compared with only 12 from U.S. companies. Miniature sensors and actuators furnish obvious benefits for Japanese industry, with its leading position in consumer electronics, robotics, microelectronics and automobiles.

The Japanese government has made the development of technologies for building small machines into a national priority. The Ministry of International Trade and Industry (MITI) last year announced funding of about $200 million over 10 years targeted toward making small actuators and sensors for fabrication of robots for pipe inspections and intelligent catheters with scalpellike cutters for surgery. The micromachine technology project, as it is known, is one of the first large MITI undertakings in which U.S. companies have been allowed to participate. As part of the program, SRI International in Menlo Park, Calif., has been commissioned by MITI to make artificial muscle fibers out of electrostatic actuators or an elastomeric material such as polyurethane.

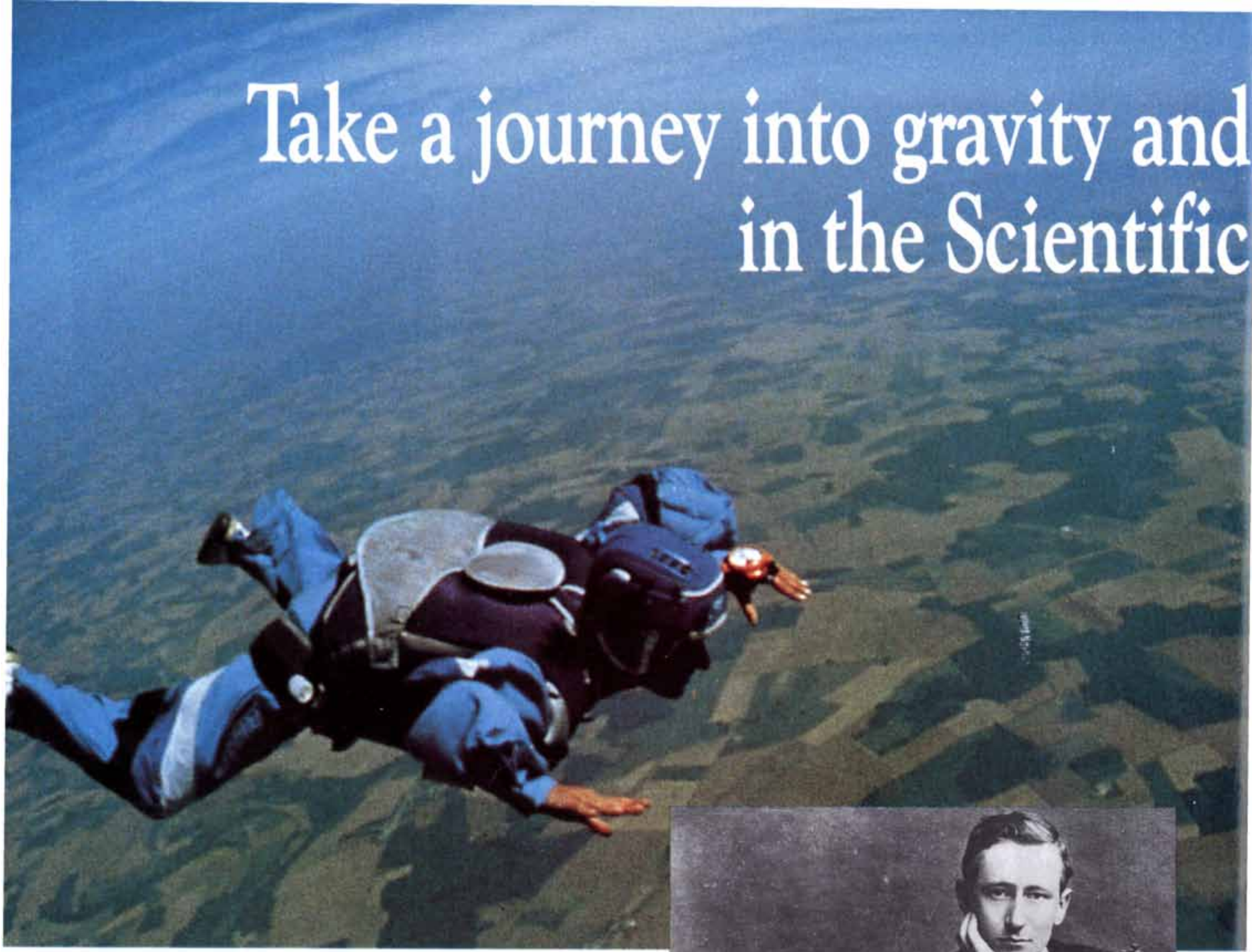
MITI has been involved in faltering negotiations for a contract on control systems for small machines with Texas Instruments. The Dallas-headquartered company has made perhaps the most noteworthy prototype of a micromechanical system: an array of two million microscopic metal mirrors on a silicon chip. Each 16-micron-square mirror tilts to an on-or-off position to control the amount of light reflected onto a display screen.

The MITI program encompasses not only lithographic mass production with silicon but a grab bag of methods to machine tiny parts from metal and plastic by cutting and grinding each individual element. Also, Sumitomo Electric Industries, the first Japanese company to exploit LIGA, will employ the technique to make small ceramic microphones for use with the pipe-inspection robots.

Although 100 companies applied to join the MITI effort, only about a quarter of that number were selected. But some prominent Japanese corporations view the project with lukewarm interest, while internally they continue to focus on silicon machining processes. A few Japanese researchers expressed the belief that MITI chose not to target silicon micromechanics to counter the perception that Japan copies and commercializes research that originated in the U.S., a source of potential trade tensions. MITI denies this contention.

Hitachi, the diversified electronics giant, is a participant, but its primary re-

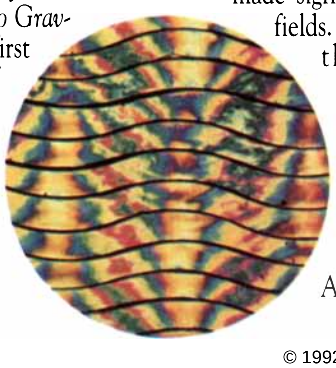
© 1992 SCIENTIFIC AMERICAN, INC



Meet the giants who make science happen. (Guglielmo Marconi in 1896. Courtesy Marconi Company Ltd.)

Now you can preview *A Journey into Gravity and Spacetime* for 15 days risk-free when you join the Scientific American Library.

Gain fascinating insights into one of nature's fundamental forces. Discover the latest thinking about Einstein's general theory of relativity. Enjoy the special viewpoint of one of our most imaginative physicists. All when you join the SCIENTIFIC AMERICAN LIBRARY and receive John Archibald Wheeler's *A Journey into Gravity and Spacetime* – the first in a remarkable series of books that will take you on a wondrous journey to the frontiers of scientific discovery.



Enter the world of force and matter. (Polarized light view of stressed graphite fiber-epoxy resin composite. Courtesy Tsu-Wei Chou.)

Your invitation to share the ideas of today's leading scientists

SCIENTIFIC AMERICAN LIBRARY books are created in a unique collaboration between our editors and renowned scientists – men and women who have made significant contributions to their fields. The result is a series of books that are readable and *authoritative*.

Beautiful, informative volumes you'll treasure for years to come

Entertaining and informative, the SCIENTIFIC AMERICAN LIBRARY provides you with an excellent resource for study and research. Each volume features abundant photos, paintings, and illustrations, most in full color. Helpful diagrams, charts, and tables clarify points made in the text.

And SCIENTIFIC AMERICAN LIBRARY books are produced to exacting standards. Printed on high-quality paper and bound in cloth, they make a handsome addition to your library, providing you with years of pleasure.

Preview a new title every other month at no risk

Your premier SCIENTIFIC AMERICAN

© 1992 SCIENTIFIC AMERICAN, INC

search thrust is still with silicon. The company is increasing its staff of four full-time micromechanics researchers to 10 and is adding a new clean room and semiconductor fabrication line for micromechanics. Hitachi has already achieved a grounding in the technology by making prototypes for acceleration sensors, which it supplies to automobile companies.

Kazuo Sato, who heads the laboratory, drills researchers in the complex nuances of the field. Every two weeks he organizes a session in which a researcher is assigned to study a paper from a recent technical conference and report on it to the group. Unlike at the university, projects are closely tied to the company's broad technology portfolio. Earlier this year Sato presented the lead paper at an international conference, in which he described a new type of microvalve that may permit precise control of the amount of gas injected into a vacuum chamber, a procedure essential to depositing single atomic layers onto a substrate for making semiconductors that operate using quantum effects.

While Hitachi builds a clean room, a U.S. pioneer in micromechanics has walked away from the field. AT&T Bell Laboratories was among the first major companies to get in—and among the first to leave. Researchers at the Holmdel, N.J., laboratory built one of the early micromotors, did research on polycrystalline silicon and constructed a "wobble" motor in which the rotor turns irregularly on its bearing, a design that reduces adverse frictional effects.

Bell Labs' micromechanics program was a victim of the much publicized decision by AT&T's research arm to narrow its once wide-ranging pursuits. In 1991 Kaigham J. Gabriel, the last specialist in micromechanics, went to another job after about five other colleagues had been transferred to other research or had left the company. Gabriel had helped stitch together the Bell Labs effort during the mid-1980s.

A year after he turned out the lights at AT&T, Gabriel was handed the reins of a government program that may give him a chance to shape the future course of micromechanics in the U.S. Gabriel's new job at the Defense Advanced Research Projects Agency (DARPA) is to nudge leading academic research toward the laboratory door. "In a sense, the honeymoon is over. People aren't willing to see another gadget of the month come out of this field," Gabriel says.

The DARPA effort is one of the largest U.S. government programs in micromechanics ever launched, even though the budget of $20 million to be expended over three years is still relatively paltry compared with estimates of spending in Japan and Germany. Gabriel, however, brings a focus to his job that the run-of-the-mill bureaucrat may not. During his tenure at AT&T, he organized many seminars to help define and explain the field to others. Even more important was the time he spent in the laboratory.

While building a micromotor at Bell Labs, Gabriel came up against the two-dimensionality of a microactuator and the difficulty of trying to extract useful forces from a machine with circular parts whose diameter is less than that of a pinhead. So instead of miniaturizing replicas of large industrial machinery, as does Henry Guckel, Gabriel tried to build actuators that could take advantage of the "flatland" produced by surface micromachining.

Post-Modern Times

Many of these ideas were executed during a year Gabriel spent on sabbatical from Bell Labs at the University of Tokyo's Institute of Industrial Science. Brainstorming with Hiroyuki Fujita, Japan's leading academic researcher in microactuators, and others turned out a series of devices that Charlie Chaplin would have been hard-pressed to recognize: membranes a few hundred microns in diameter inflating above and then deflating back to the chip's surface like tiny blowfish and matchstick-like parallelograms that expand and contract at their vertices. (On his own, Fujita has produced actuators that curl up off the surface and back, just as cilia do.) An army of parallelograms arrayed across a chip's surface might produce enough force and movement to manipulate larger structures. "The individual actuators are stupid," Gabriel comments. "But together they behave like a complex device."

The parallelograms were just the kind of moving parts that Gabriel had hoped Bell Labs might consider for switching an optical signal mechanically, forgoing the need to convert it into an electronic impulse and then back into its optical form, as is done now. Gabriel believes micromechanics is "something they're going to have to get back into." Indeed, Gabriel returned to Bell Labs in late summer to make a presentation on the DARPA program.

What he described to his former employer is a program that will push researchers to translate years spent in refining microfabrication expertise into practical machines or sensors. Among the contracts:

• The most sensitive acceleration sensor built to date. Cornell Universi-

© 1992 SCIENTIFIC AMERICAN, INC

ty has been refining for more than six years a reliable process for making the notoriously hard-to-manufacture sensing elements and actuators for scanning probe microscopes used to write company names atom by atom. If all goes well, the research team there will use arrays of sensing tips to detect movements in any axis caused by forces ranging from a few millionths up to several tens of times that of gravity. Such a device could become the fundamental technology for a pocket version of the inertial navigation systems that allow pilots to know where they are anywhere on the planet.

• A wristwatch-size device made up of several chips that will measure a range of variables (barometric pressure, temperature, humidity, geographic position and the presence of gases such as carbon dioxide) that could have broad application in health care, industrial control and remote environmental monitoring. More important, this University of Michigan project could become a prelude to a new approach to computerized design. A gas sensor or a pump will become just another design element in the software toolbox alongside the transistor and the capacitor.

• An actuator from Stanford University whose motion would be perpendicular to the surface of the chip. Such upright digits could converge to grasp biological tissue or move small objects in sequence as if they were riding on an assembly line.

As part of another program, DARPA is supporting an experiment in which designs of surface-micromachined structures from all over the U.S. will be processed together at a semiconductor facility in North Carolina. This trial will probe whether the field is mature enough to support a facility for entrepreneurs and academics to order small samples of micromechanical systems for a nominal fee, similar to a DARPA program for the microelectronics industry. As the field rises on the foundation established by microelectronics, progress may come relatively quickly. "We don't have to build an infrastructure," Gabriel comments.

That may not quell the skeptics in the audience. On a recent trip to Germany, Henry Guckel, the LIGA proponent, heard the inevitable question about what his mote motors were good for. Guckel urged patience, explaining to the assembled industrialists and academicians that they could take delivery of micromachines now or later. Either way, they might be useful. But it might be better to wait, Guckel told the crowd. "It depends what you want to drive, a Yugo or a Mercedes."

© 1992 SCIENTIFIC AMERICAN, INC

Building Networks

New York Telephone rethinks work to regain lost customers

Leroy Gilchrist knows how it feels to be thoroughly versed in a job—and to be good at it. A 22-year veteran of the New York Telephone company and union member, Gilchrist has repaired and spliced every kind of telephone line from twisted copper pairs to optical fibers in buildings throughout New York City. "You get into a mindset," he says. "You say, 'This is the way I've been working. So I don't want to hear anything about changing it.'"

Yet for the past year, Gilchrist has been part of an eight-man commando team charged with doing just that: rethinking how the telephone company goes about "provisioning," or providing customers with one of its most advanced products, and then radically restructuring how the work gets done.

At stake are hundreds of millions of dollars of business. Since the deregulation of the telecommunications industry in the mid-1980s, New York Telephone's share of the Manhattan market for privately leased, high-speed data highways, called T1 networks, slipped from 100 percent to about 64 percent this year. These services are used by businesses from banks to long-distance communications companies to ship 1.5 million bits of data per second between offices along a single, private path. Analysts estimate that the U.S. market for leasing such lines tops $2 billion. "We lost significant market share in one of the premier products in our arsenal," concedes Douglas J. Mello, a group vice president at New York Telephone.

Now Mello is betting that Gilchrist and his collaborators have designed a plan for reversing that slide. "There's no reason why we shouldn't do better," Mello insists. Seventy-five to 80 percent market share would be "reasonable," he suggests. What is more, Mello hopes that the technique of rethinking work practices will be applied more broadly within the telephone company.

The effort puts New York Telephone at the forefront of a sweeping trend that promises to be as embracing as the 1980s doctrine of quality. From the halls of the Massachusetts Institute of Technology to the conference rooms of boutique consultancies, management gurus are telling corporations to take a long, hard look at how work really gets done and fix the process.

MEMBERS of the telephone company's T1 Center, led by Eric Wilson (seated, left), *are at the heart of an experiment in redesigning work practices. Photo: Dan Wagner.*

Such concerns hit home at New York Telephone's parent, Nynex, in early 1991 as its share of T1 business continued to melt away. Mello toyed with hiring consultants. "But you only really get discernible change if the people closest to the problem are involved," he says. So, in a first for the telephone company, Mello appointed a design team staffed jointly by "craft," or union, workers and first-line managers and set them to the task of sharpening the way the company handled orders for T1 lines. When he learned that researchers in the Nynex Science and Technology expert system laboratory were already beginning to model work processes, Mello pulled them into the project as "facilitators."

Among the researchers was one consultant, anthropologist Patricia Sachs. "Yes, we get a few smirks about the anthropologist," Mello says. "But the business is about people, and we had to understand how the job gets done."

Gilchrist and his teammates found a multitude of reasons why the phone company had stumbled—126 to be precise. "We looked at the process from A to Z," Gilchrist says, from the central office through the field construction. "I think it was the first time that anybody had an idea of what the whole process looked like," he adds. Filling an order, the group discovered, involved 126 worksteps and more than 40 people.

Although many individuals were responsible for some part of a T1 installation, no one saw the job through until the end. For instance, customers spoke first with a service representative. The "rep" would take down some details to start the order and suggest when the company might hook up service.

But that timetable was only a best guess—the rep was unlikely to know any of the engineering details that would determine how quickly service could be connected. And although the rep would have recorded all the data needed to process the order, engineers would likely need other information—a guarantee of future delays. "We used to think, 'If only those other guys could have done their job right, then there would be no

© 1992 SCIENTIFIC AMERICAN, INC

problem!'" recalls Eric Wilson, a member of the design team and a former rep. "But we never really knew what was going on at that other end."

Orders from the sales reps were then routed to the Digital Facilities Quality Center, which passed jobs, like so many soccer balls, to service supervisors, who in turn routed them to various technicians. Scheduling was coordinated by a computer, which distributed job tickets. Should one technician encounter a snag, he might simply go on to the next job as another worker was sent out to solve the problem. "So a customer might see three dispatchers marching in and out and still not have T1 service," Gilchrist points out. And only rarely did the reps—who talked with customers—ever talk with engineers or with technicians.

With the help of the Science and Technology facilitators, Mello's group began to probe the work practices more deeply, modeling the work flows on a computer. Some simple changes seemed obvious: the design team quickly recommended rewriting the form filled out by the reps as they took orders to include more appropriate information. "This sounds simple," says Elizabeth R. Graham, who headed the facilitators' group. But it took almost a year to elicit and coordinate all the relevant questions for the new form, she adds.

More fundamentally, the team explored paradigms, or descriptions, of how different groups interact. Led by Sachs, the members talked about how African bands function, how the telephone company used to work and how they would organize a start-up company that sold T1 services. They visited other divisions of the company and even a branch of Corning, Inc., to watch how self-managed teams performed.

The group even tried imagining the perfect employee—"Melvin"—who could do everything himself, then gradually gave Melvin perfectly adept assistants. By thinking about what information Melvin needed to get his job done, Sachs notes, it became clear who needed to talk to whom to move ahead a job like setting up a T1 line. "You thought you understood your job, but you were only acting in a little box," Wilson observes. "We came out of that box," Gilchrist adds, "but reluctantly."

Over the weeks the design team articulated a set of guiding principles, or ways that the team members wanted to work. Ensuring that people were responsible for an entire project—from order to installation—was on the list. Also important, Sachs adds, was the belief that knowledgeable workers are valuable assets. The practical implication of that principle is that computer systems should be designed to help people learn to do their jobs rather than simply taking over tasks such as scheduling.

Within six months, the design team was ready to offer senior managers its new approach. "We wanted four organizations collapsed into one," Graham recounts, and proposed shifting responsibilities so that one small group would be responsible for an entire T1 job. Coordinating the work would be a "T1 Center." Most daring was a proposal to eventually create a job that would bridge the traditional union-management divide. After the presentation, Sachs recalls, "someone said, 'We asked you to organize a radical change, and by God, you did it!'"

Transforming the ideas into action turned out to be harder than the team expected. "We had a whole lot of good stuff," Gilchrist says, "but the only people who knew it were us. And wow—the rest of the world was big!" Senior management decided to give the new plan a six months' trial in downtown Manhattan but required one compromise: rather than try to negotiate a new job title, New York Telephone would simply bring three types of workers—reps, engineers and trunk assignors—into the T1 Center and sit them down side by side. "They'll do all the front-end work," Graham explains, and can call on one another to handle questions.

From the T1 Center, the orders will flow to a turf coordinator, responsible for some section of the city. That coordinator will track the progress of the job until completion, keeping in close touch with the technicians. According to the simulations run by the Science and Technology researchers, the plan should enable a team of a dozen—rather than 40—people to set up a new T1 service in two or three days, down from more than seven. The coordination relies on communication among people, rather than on computer scheduling, points out David M. Torok, a facilitator.

Late this September the trial was scheduled to swing into action. Mello will be watching for evidence that customers are happier and that the time for setting up services has dropped. The Science and Technology researchers will be monitoring how people learn to collaborate and looking for clues about how to design future software systems. Wilson, who will be acting manager of the T1 Center, plans to urge workers to keep seeking ways to smooth the process. Gilchrist has returned to his work as a splicer and is already helping people sort out the new teamwork approach. "It's sure been a real education," he says with a smile. *—Elizabeth Corcoran*

Cool Sounds

An acoustic compressor pumps up the volume on refrigeration

After years spent developing whisper-quiet refrigerators, appliance makers may turn to noise to cool the next generation of iceboxes. A physicist-turned-entrepreneur has developed a lubricant-free sonic compressor that promises to be more energy efficient than standard compressors and can use refrigerants not based on ozone-depleting chlorofluorocarbons (CFCs).

Timothy S. Lucas, president of Sonic Compressors Systems in Glen Allen, Va., believes his compressor may prove to be the tonic the appliance industry needs to meet the upcoming stringent CFC regulations and energy efficiency requirements. Lucas says his start-up company has built working prototypes and is negotiating a contract with a major refrigerator maker. He hopes to demonstrate his device in a domestic refrigerator in early 1993.

What makes the instrument promising is that it is a drop-in replacement for conventional compressors and so will not require retooling or other expensive procedures. In addition, "we should be able to make the compressors at a comparable price," Lucas says. Coupled with a projected improvement in energy efficiency of 30 to 40 percent over existing compressors, manufacturers may save on more expensive techniques for efficiency, such as increasing insulation or adding larger heat exchangers and more powerful fans.

Standard compressors in refrigerators generally rely on pistons or rotors to compress the cooling gas. Because refrigerators are expected to function trouble free for 15 to 20 years, the compressors must be lubricated to prevent excessive wear. And the refrigerant and the lubricant must be compatible. The most ideal mix has been mineral oils and CFC 12, which is scheduled to be phased out over the next few years, possibly by 1996. As a replacement, the chemical industry is pushing hydrofluorocarbon (HFC) 134a, a substitute that developers claim has no known effect on the ozone layer and should produce only a minimal greenhouse effect.

Some refrigerator engineers worry that lubricants compatible with the new refrigerant may not be developed in time. The appliance industry has been evaluating ester-based oils, but tests so far have not been encouraging. "The lubrication properties are not as good as those of mineral oils," says Carl Offutt, the general manager of engineering at

© 1992 SCIENTIFIC AMERICAN, INC

the Whirlpool Corporation. Furthermore, "funny things start happening when you add other compounds to improve the lubricant's effectiveness," he remarks. Over time the ester oils react with HFC 134a and other oils used to manufacture components to form a waxy residue that plugs up the refrigerator tubes. The oils, too, may react with water to corrode the internal parts, so that refrigerators would have to be made in a moisture-free environment.

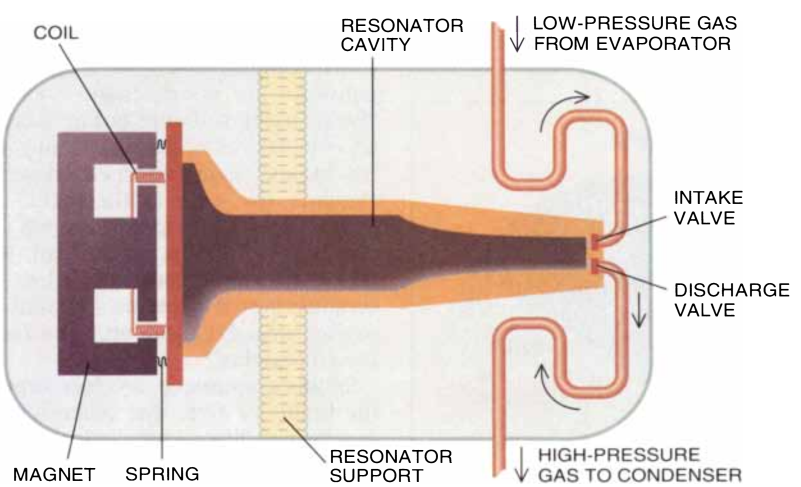
In contrast, Lucas's compressor needs no lubrication, and that means "throwing away oil restraints" in finding an environmentally friendly refrigerant, Lucas says. The sonic compressor is simply an oddly shaped tube that acts as a resonance cavity for the refrigerant. The entire resonator moves back and forth about 50 microns along its cylindrical axis at about 340 hertz. The oscillation creates a standing wave in the cavity. Because the cavity is designed so that the standing wave reinforces itself, the pressure changes achieved in the tube are large. In terms of sound pressure, the amplitude is about 200 decibels, but the compressor is nonetheless quiet because the mass of the tube prevents any sound from escaping. A valve vents the pressurized refrigerant into tubing that circulates the fluid.

The biggest problem Lucas faced in his design was eliminating shock waves that formed in the cavity, wasting power as heat and thereby limiting compression. Lucas spent a year at Los Alamos National Laboratory working with acoustic and engine-cycle expert Gregory W. Swift to find a solution. "The trick was the geometric design of the resonator," Lucas points out. The shape made the higher harmonics that caused the shock waves to interfere with one another, leaving only the fundamental frequency in the cavity.

To an industry known for extreme caution in introducing new technology, a radically redesigned compressor may seem risky. Nor have manufacturers forgotten General Electric's debacle with its rotary compressor in the 1980s. After rushing it into production, GE discovered the compressor was defective; the company spent an estimated $450 million to recall the products. But Lucas is unperturbed. "It appeared to be more of a management problem rather than an engineering one," he notes.

Refrigerator companies may have few other alternatives if they cannot meet tough energy efficiency standards cost-effectively or find an HFC 134a–compatible oil. Manufacturers should be able to meet 1993 efficiency guidelines by tweaking existing parts. But the oil compatibility issue remains, well, sticky. One industry expert thinks manufacturers could end up shortening the warranted lifetimes of the machinery if the gumming is not solved. All the more reason, perhaps, to chill out with a new compressor. *—Philip Yam*

Refrigerating by Sound



The sonic compressor is driven like an ordinary loudspeaker. The oscillation creates in the resonator cavity a high-amplitude standing wave, which compresses the refrigerant gas. A one-way valve vents the gas to a condenser, which liquefies and cools it. The liquid circulates around the space to be cooled, evaporating and thus absorbing heat. The gas is then pulled back into the compressor to repeat the cycle.

Semiready

A consortium prepares for the second half of the battle

Murphy's second corollary says everything takes longer than expected. Saving the U.S. semiconductor industry is no exception.

In 1987, with nerves still jangling from watching Japanese competitors gobble up the market for dynamic random-access memories (DRAMs), U.S. industry and government created Sematech, the Semiconductor Manufacturing Technology research consortium. According to the founders, Sematech would develop manufacturing techniques that any U.S. chip maker could use, thereby honing America's overall competitive edge. To pay for the project, industry and government each agreed to spend $100 million annually for five years.

That time is up. Although U.S. memory chip makers are not yet on firm ground, many believe Sematech has instead helped stabilize the other half of the chip business—namely, the semiconductor manufacturing equipment industry. Now, supporters suggest, Sematech can turn back to its original mission of directly addressing the needs of chip makers. As a result, Sematech's leaders are pounding the pavement at Capitol Hill, arguing that they have spent the money wisely and that the government should continue the consortium's allowance. "I don't think there is anyone who will not say Sematech has done a lot of good things for U.S. industry," says Richard A. Aurelio, an executive vice president with Varian Associates in Palo Alto, Calif.

Representatives of the Bush administration are also enthusiastic. "We feel that Sematech's been extremely effective," says Arati Prabhakar, who directs the microelectronics technology office at the Defense Advanced Research Projects Agency (DARPA). Yet as Sematech has reached the five-year mark, DARPA's obligation is over, she says. "We're not in the business of perpetuating block funding," Prabhakar adds. The government has proposed trimming funding to $80 million in fiscal 1993. (Observers nonetheless believe Congress will restore Sematech's $100 million.)

Measuring Sematech's performance precisely is tricky. Earlier this year market research analysts estimated that the erosion of American firms' market shares in chips and equipment had stopped; some reports indicated that at least the top few equipment makers gained some market share during 1991.

But industry leaders caution that any

© 1992 SCIENTIFIC AMERICAN, INC

such gains are short-lived at best. For the equipment industry, at least, "the claims that we're regaining market share against the Japanese aren't true," Aurelio says. Japanese chip makers bought little equipment last year, he points out. "As soon as the Japanese start investing again, we'll be down."

Even so, equipment makers say the program has enabled them to understand better their customers' needs. Since they were brought together under the Sematech umbrella, tool and chip producers speak more freely. Although individual firms were working on the quality and reliability of their products, the only systematic cross-industry program that took hold was Sematech's Partnering for Total Quality, observes Papken Der Torossian, chairman of the Silicon Valley Group in San Jose, Calif.

Tool and chip researchers also collaborated on standard techniques for approximating the mean time between failures and for estimating the cost of a new tool, says Neil R. Bonke, president of General Signal's semiconductor equipment operations in Santa Clara, Calif. "Before Sematech, buyers wouldn't accept your figures," he explains. They relied instead on homegrown models for calculating likely costs and failure rates. "Now when asked, we're all using the same sheets of music."

Sematech also helped the industry develop specific tools. Even though Applied Materials is the top U.S. semiconductor equipment maker, "we have huge R&D expenses. It is difficult to go it alone," says Dan Maydan, executive vice president at Applied Materials in Santa Clara, Calif. Technical—and financial—aid from Sematech accelerated the development of a new etching tool.

Support from Sematech enabled General Signal and the Silicon Valley Group to buy out faltering U.S. toolmakers and have another go at building advanced lithographic steppers, the devices that project integrated-circuit patterns onto semiconductor wafers. The Silicon Valley Group received significant aid from IBM, in the form of a pledge to buy 20 advanced steppers. Yet without Sematech the deal would have fallen through, Der Torossian says.

In its sixth year, Sematech's leaders aim to refocus the consortium's attention on issues of immediate concern to chip makers—such as computer-integrated manufacturing and fighting contamination in clean rooms. At the same time, Sematech plans to help chip manufacturers advance to ever more sophisticated and densely patterned chips.

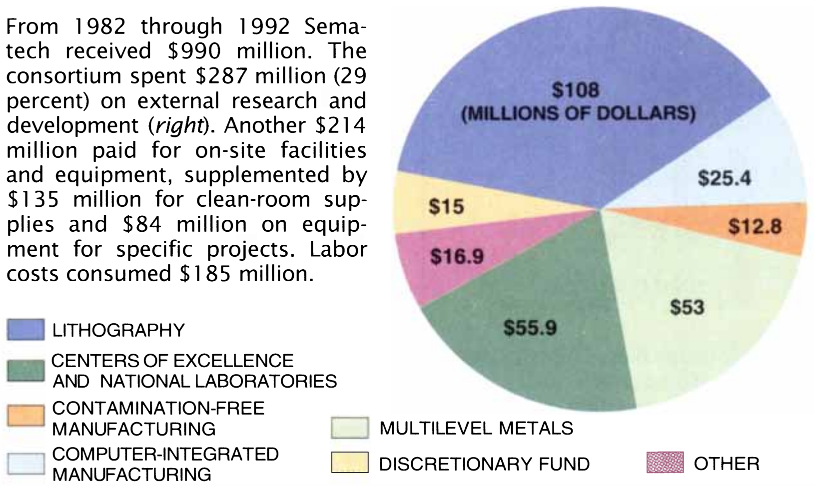
Equipment makers worry they will be left behind. More complex chips, Bonke observes, will require new tools. "I'd like to see more direct investment in the form of guaranteed loans to manufacturers to develop new technology," Aurelio adds.

Others suggest there are more roles for Sematech. Der Torossian would like to see Sematech serve as a clearinghouse for information about all the government-funded research programs in microelectronics. Sematech "should be used as a catalyst to organize and maximize the semiconductor R&D that we do in the national labs," he says.

For now, Sematech has its eye on securing the funding it believes it needs to move chip manufacturing to the next generation of devices. The priority of Sematech's first five years was to catch up with Japanese competitors, says DARPA's Prabhakar. "Now," she adds, "the issue is to provoke fundamental change." *—Elizabeth Corcoran*

Where Sematech Spent Its Money

From 1982 through 1992 Sematech received $990 million. The consortium spent $287 million (29 percent) on external research and development (*right*). Another $214 million paid for on-site facilities and equipment, supplemented by $135 million for clean-room supplies and $84 million on equipment for specific projects. Labor costs consumed $185 million.



SOURCE: General Accounting Office

Intercepted Messages

New biotechnology drugs target intracellular communication

The recognition that cells receive instructions from hormones and growth factors provided a foundation for the biotechnology industry in the 1970s. At the time, few people categorized proteins such as insulin and erythropoietin as "first messengers." It was enough to know that such proteins instigate cellular change by binding to receptors in the membrane surrounding cells—and that making them in quantity would provide treatments for diabetes and anemia.

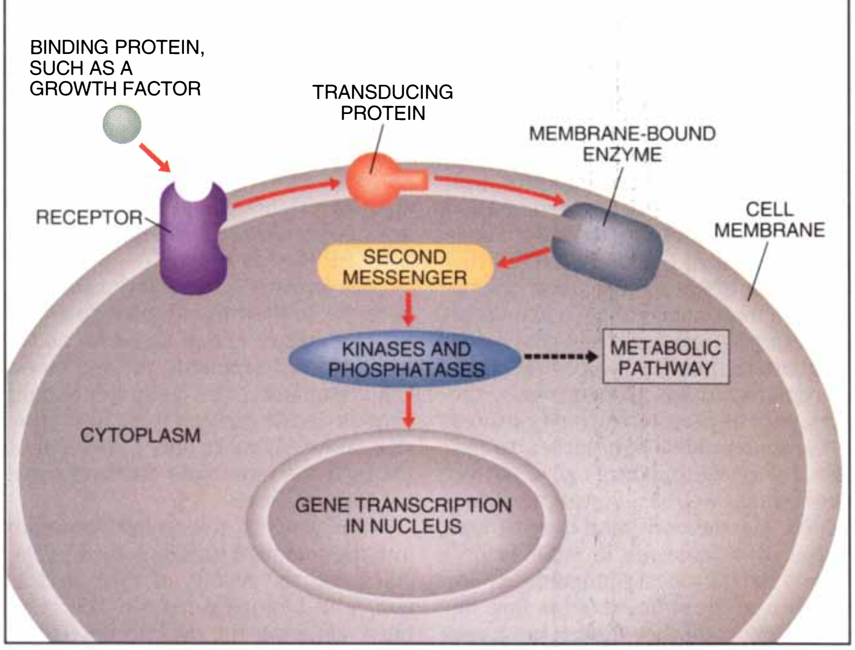
Today the way that communication proceeds inside the cell after a surface receptor is contacted is the rallying interest of a new group of biotechnology companies. When a receptor is bound, its configuration shifts. That action summons a variety of so-called second messengers, which in turn signal other chemical carriers inside the cell. This "psst, pass it on" process, known as signal transduction, is opening pathways for treating asthma, allergies, arthritis, cancer, cardiovascular disease, psoriasis and other disorders. "It is now widely accepted that many diseases are the result of dysfunction in signal transduction pathways," points out Arthur G. Altschul, Jr., director of planning for Sugen in Redwood City, Calif.

Firms including Ariad, Cadus, Onyx, Sphinx and Sugen intend to manufacture small molecules that block or mimic the action of second messengers and their followers. All believe that signal transduction offers numerous points for pharmaceutical intervention, even if the pathways are poorly marked as yet. "We're trying to figure out an incredibly complex railroad and switching system by just having direct experimental evidence for some of the parts," explains Glenn L. Cooper, executive vice president of Sphinx in Durham, N.C. "How the conductors talk, what the common switch boxes are and what the overall timetable is are only now being pieced together."

Sphinx's approach revolves around the lipids, or fats, that make up cell membranes. "It's a particularly attractive place to work," Cooper says, noting that lipids regulate certain key second messengers, such as protein kinase C (PKC), an enzyme that mediates cell functions, including inflammation and proliferation. When a growth factor binds to a receptor on the cell surface, a lipid molecule lodged in the membrane is chemically cleaved in two. One of the frag-

© 1992 SCIENTIFIC AMERICAN, INC

Steps in Intracellular Signal Transduction



SOURCE: Ariad Pharmaceuticals

ments is released, activating PKC inside the cell and kicking off a cascade of cell division and inflammation. "That's when all hell breaks loose," Cooper declares.

Sphinx believes blocking the signal that switches on PKC may prove beneficial in treating a spectrum of illnesses, including psoriasis, a chronic skin disorder that affects 2.5 million people in the U.S. This October psoriasis patients at the University of California at Irvine were scheduled to begin testing an ointment that interferes with the action of PKC.

Meanwhile Ariad Pharmaceuticals is exploring drugs for treating allergies and asthma. Because patients with these illnesses experience acute attacks, "you can test a drug and know whether it works in 15 minutes," observes Harvey Berger, the Cambridge, Mass., company's chief executive officer. In contrast, cancer drugs usually require lengthy trials in large numbers of subjects. The firm began corporate life earlier this year with a private placement of $46 million, the largest yet garnered by a biotechnology company.

Ariad's initial target is an interaction that results in the release of histamine—a phenomenon all too familiar to allergy and asthma sufferers, who experience the itchy, irritating results. When an allergen such as pollen enters the body, the immune system dispatches an antibody to meet it. The complex then binds to a receptor waiting on the surface of a mast cell. The binding alters the receptor's intracellular configuration, an event that prompts the cell to spew out its stores of histamine. "It's like a bolt of lightning hitting a tree and causing the roots to change," says Joan S. Brugge, Ariad's scientific director.

A drug to prevent the altered roots of the mast cell receptor from contacting the intracellular components that trigger release of histamine might be effective against allergies. Ariad says it intends to begin testing such a compound in humans within several years. The company expects that effort to help it devise other drugs. "What we learn from mast cells will be relevant to other actors of the immune system, such as *T* and *B* cells," Berger observes. "There are a lot of similarities already" that provide insights into rheumatoid arthritis and other diseases, he notes.

Other researchers are interested in how cells pass messages by transferring electrically charged groups of molecules from one protein to another. Second messengers often spur these movements, which are carried out by specific enzymes. For example, kinases add negatively charged phosphate groups; phosphatases reverse that action.

In various kinds of cancer, and possibly non-insulin-dependent diabetes, the culprit appears to be the enzyme tyrosine kinase. "It's a gas pedal for the specific functions of cell growth and glucose uptake," says Altschul of Sugen. The enzyme's ability to transfer phosphate groups is critical to the way growth messages echo through cells. Altschul notes that phosphatases can be thought of as brake pedals.

At Sugen, researchers are screening diseased tissues to see if they find receptors that are not present in or are somehow different from those in normal tissue. They presume these receptors are a mechanism for the disease. If the receptors contain tyrosine kinases, Sugen begins looking for matching phosphatases or other means of turning off the process. The company expects to begin testing drugs based on these discoveries in humans in two years.

Following a similar vein of inquiry, Cadus in New York City is seeking to inhibit the transfer of molecular groups that alter receptors coupled to so-called G proteins. The receptors for light in the back of the eye act on G proteins, explains Samuel D. Waksal, chief executive officer of Cadus and its parent company, ImClone. "In the eye, one photon of light activates a receptor. Then G-binding proteins set off a cascade of messages that is amplified until you see the light," he notes. Inflammation and allergy and many other responses to single molecules may proceed the same way, Waksal says. But chronic diseases can develop when a signal does not stop. Cadus is intent on blocking transfer events known as methylation and prenylation that enable G proteins to couple to receptors.

There are many regulatory proteins hard-wired into the pathway from cell membrane to nucleus, reminds Frank McCormick, vice president of Onyx. The Richmond, Calif., spin-off of Chiron is devoted to designing cancer drugs based on signal transduction. "We're going for the *ras* pathway, which includes tyrosine and other kinases," he says. Mutations in the *ras* gene (which is involved in the growth of normal cells) lock it into an active state in many cancers, causing signal transduction pathways to be hyperactive. "Anti-*ras* drugs are likely to slow down the growth of tumors—for how long is unknown," McCormick speculates. "Maybe the cells will die out or will sit latent while the drug is present."

McCormick acknowledges that shutting down pathways used by normal cells raises concern about potential ill effects such drugs might have on healthy tissues. But experiments in mice with a variety of cancers indicate that removing the genes responsible for destructive pathways has not harmed the creatures. "It seems that normal cells find a way around," he says. The early promising results from signal transduction research are encouraging scientists to follow pathways they might otherwise have avoided. —*Deborah Erickson*

© 1992 SCIENTIFIC AMERICAN, INC

Intuitive Design

Artificial intelligence helps a drugmaker learn what works

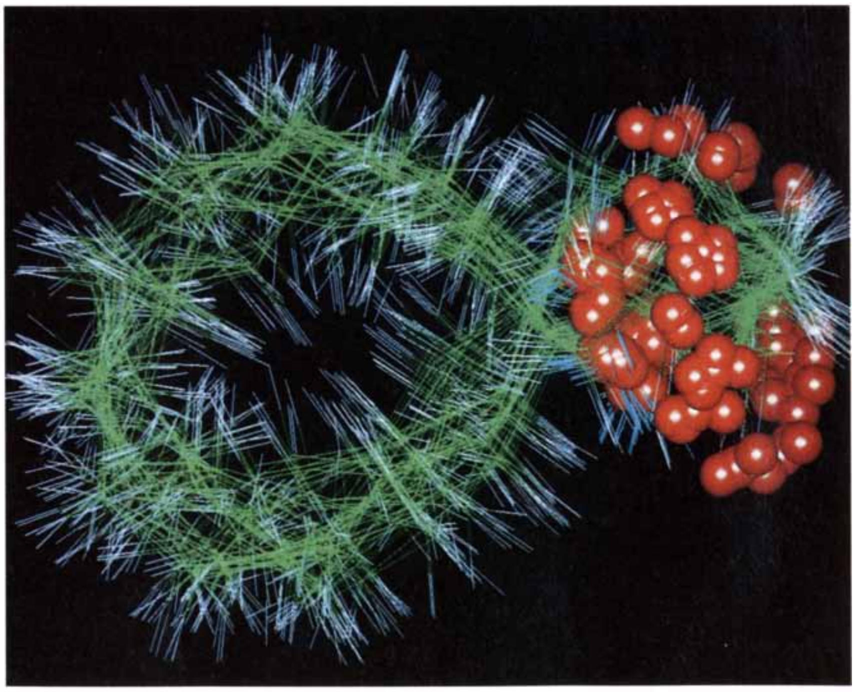
If at first you don't succeed, then try, try again: such is the tradition of drug development. The sometimes intuitive, basically trial-and-error process typically involves screening 20,000 to 40,000 compounds to find a single promising lead. For every 4,000 leads examined, often no more than one becomes a marketable drug. To rein in this randomness, medicinal chemists have been attempting for the past 10 years to construct drugs that fit precise targets, such as receptors on cell surfaces.

Computer-aided approaches to "rationally" designing drugs have not yet worked as well as researchers had hoped. The methods tend to rely on an image of a molecule in what is presumed to be its ideal configuration—for instance, at the moment of binding. Because few computer design programs can acknowledge the mutual shape changes between a molecule and its receptor, drug designers may be misled into creating a "perfect" compound that cannot function in the body.

Arris Pharmaceuticals believes artificial-intelligence (AI) systems are a better way of utilizing the strength of computers in drug discovery. Computer algorithms similar to the kind that enable U.S. military cruise missiles to find their targets can also be written to recognize the shapes of molecules, explains Michael J. Ross, president and chief executive officer of the South San Francisco firm. Similarly, programs developed by robotics researchers to move a mechanical arm without knocking down obstacles allow the computer to contemplate how a drug contacts its receptor. "We're not developing expert systems to replace medicinal chemists," Ross notes. "We're developing tools to augment their intuition."

The company is using its AI systems in-house, initially to optimize drugs for asthma and to create an orally active form of erythropoietin, the red blood cell stimulator. "The computer should help us decide early on what to synthesize, so we will make only 10 percent of the molecules we would otherwise have to," Ross declares.

In the process, researchers program into the computer descriptions of molecules and the results of their performance in biological assays. The computer analyzes the data to determine what all the compounds that work have in common and how they differ from those that do not work. The scientists apply the computer's observations in the next round of experiments, steadily building a data base of experience. "People just cannot remember all the data they've seen," says Tomás Lozano-Pérez, a researcher in computer vision and robotics at the Massachusetts Institute of Technology and an Arris consultant.



MOLECULAR SURFACES, analyzed in detail with artificial intelligence, may prove to have patterns that influence the performance of drugs. Photo: Arris Pharmaceuticals.

© 1992 SCIENTIFIC AMERICAN, INC

"Machine learning algorithms will find patterns that exist in all the data."

By relying on computers to track experiments, Arris hopes to be able to model molecules in much greater detail than is now common. Most computer representations of drugs focus on a few points and the distance and angles between them, Lozano-Pérez explains: "We're finding you need more like 100 data points to characterize a molecule properly." Attributes such as molecular charge and hydrophobicity add dimensionality that is difficult for humans but simple for computers to consider.

Arris scientists anticipate that mapping the surfaces of molecules in detail will free them of established notions concerning the way certain classes of molecules are influenced by structural features such as "backbones." Ross notes, "It's not just how much a surface sticks out in one direction but also what's in between that counts."

A key aspect of the Arris approach is that it takes advantage of negative data. "If you're hiking up a mountain, you'd like to know where a mudslide has closed off a path, so you won't bother going that way. You'd at least like to mark your map for next time," reasons Richard H. Lathrop, senior AI scientist at Arris. The system he and the computational group at Arris are creating will record the topography of molecules. "It will tell you, whatever you do, don't have something pointy sticking out of the left side," he says.

The company believes it will be able to apply what it learns from one class of molecules, such as peptides, to other compounds, such as small organic drugs that can be taken orally. "I think Arris is going to make money," declares Larry Hunter, director of the machine learning group at the National Library of Medicine. "Small increments of improvement lead to bigger and bigger advantages, as the Japanese have demonstrated with cars," he notes.

Peter S. Kim, a structural biologist at the Howard Hughes Medical Institute at the Whitehead Institute, observes: "It's not trivial to bring the artificial-intelligence people to the biology people and have them all mix—that's starting to happen at Arris."

"If what we have is as good as we think it is, we won't be able to stop people from following us," Ross says. Arris has attracted the top academic researchers in machine learning and pattern recognition, but competitors could still tap a rich source of highly skilled people. In these post–cold war days, classified weapons designers may find themselves invited to join a new technology explosion. —*Deborah Erickson*

© 1992 SCIENTIFIC AMERICAN, INC

A Risk Worth Taking

Poor people have few assets. As a result, few financial institutions are eager to lend them money. According to conventional economic wisdom, poor borrowers lack the income to pay off loans. Worse, if they default there is no collateral to seize in lieu of repayment, although loan sharks may circumvent this problem by seizing borrowers instead of their assets. That no one lends to the poor is generally accepted as proof there is no money to be made at it.

Several banks and financial groups are successfully abandoning this prejudice. Many poor people, it turns out, are more than creditworthy under the right circumstances. Consequently, social betterment can be profitable—both in the Third World and the U.S. "There really is a market niche for these innovative forms of lending," says Michael Carter of the University of Wisconsin. "Capital markets tend not to take care of people very well."

The programs that work rely on small loans, short repayment periods and, frequently, group lending. The most famous is the Grameen (or "village") Bank of Bangladesh, whose average loans of $67, roughly equivalent to half a year's income, must be paid back in one year. Grameen reports that 98 percent of its loans are repaid—as opposed to the country's average rate of 30 to 40 percent.

Economist Muhammad Yunus founded Grameen Bank in 1976 after he determined that lack of capital was the primary obstacle to productive self-employment among the poor. Officially established in 1983, the bank now has some 980 branches and 1.2 million borrowers. Ninety-two percent of the clients are women who sought to start their own businesses. Villages where the bank has lent money have registered improvements in education, health care and women's status.

Grameen's strategy for success is peer lending. To obtain a loan, an individual must band together with four neighbors. The group meets with a loan officer and then chooses one or two of the five to be eligible for an initial loan. Before another group member can receive a loan, the first borrowers must make regular repayments. All loans must be repaid before anyone becomes eligible for a second, larger loan.

Peer lending solves several problems that are inherent in all capital markets but are particularly thorny for the poor. The first is lack of information. If a bank does not know a client's history, it cannot adequately assess the risk of lending. Although the expected profit on a $100-million corporate line of credit can justify detailed fact-finding, the interest on $67 for a year would barely cover the telephone call for a computerized credit report—assuming that a landless Bangladeshi farmer had a documented credit history.

In contrast, people from a village or neighborhood know one another. In effect, they eliminate the costs that the lender would otherwise have to pay to determine creditworthiness. "It is hard for people in the hierarchy to have as good information as peers do," notes Joseph E. Stiglitz of Stanford University.

People considered "poor credit risks" have loan repayment rates of nearly 100 percent.

In addition, Stiglitz says, "peer monitoring addresses the theory of moral hazard." According to this theory, borrowers will repay loans to banks only if they have something to lose that is more valuable than what they could gain by keeping the cash. When the last recession hit the U.S., for example, thousands of home owners abandoned houses that were suddenly worth less than the balance on their mortgages. When people's fates are linked, however, they can create responsible peer groups and police one another. "You are tied to the group as a whole," Carter explains.

To some, the model of the Grameen Bank would not seem to be one that could be successfully transplanted to the U.S. "A lot of Americans would balk at it," Stiglitz says. Indeed, some attempts at peer lending in the U.S. have failed, perhaps in part because the neighborhoods are not as homogeneous as they are in Bangladesh. Nevertheless, more than 20 U.S. organizations use group lending to fund small enterprises. Some, such as the Women's Self-Employment Project in Chicago, have repayment rates of nearly 100 percent. (The U.S. student loan default rate is currently 19 percent.)

Despite the outstanding repayment records of these small lending organizations, some economists question whether they are really profitable. Grameen, for example, still requires a subsidy to meet its operation costs.

Yet Grameen "wouldn't take much to become profitable," argues J. D. Von Pischke of the World Bank. Alexander M. Counts of RESULTS, a nonprofit group working on world hunger, contends the bank will be in the black in a few years. "It was not profitable because it was expanding rapidly until recently," he explains. Bank Rakyat Indonesia, which provides rural credit nationwide, is already profitable, notes Marguerite S. Robinson of the Harvard Institute for International Development.

The apparent success of Grameen, Rakyat and similar institutions also exposes economists' disregard of the informal sector (small-scale self-employment) as an engine for development. As much as half of the gross domestic product of Latin American countries, for example, comes from such activities, says Gabriela Romanow of ACCION International, a nonprofit organization that has granted microloans since 1973. Meanwhile traditional economists focus on—and the World Bank tends to fund—natural resource exploitation and infrastructure development, such as road building.

Nor is encouragement of the informal sector a development strategy suited only to Third World countries. In the U.S., more than 100 organizations currently assist microenterprises by providing training, grants or loans. Congress gave the Small Business Administration $15 million this year for 35 microloan demonstration projects. "Some banks are beginning to see this as a way to broaden their normal operating procedure," says Beverly Smith of the Association for Enterprise Opportunity.

To proponents of Grameen, that barely begins to tap the possibilities. "There is no reason we couldn't have a thousand Grameens in this country," Von Pischke says.

—Marguerite Holloway and Paul Wallich

© 1992 SCIENTIFIC AMERICAN, INC

Now schools without French teachers can still have lessons in French.

Students at three high schools in northwestern suburbs of Chicago now can take classes their schools might not have been able to afford otherwise—thanks to a system provided by Ameritech.

Called Distance Learning, the system is a new application for video conferencing, a service originally intended for business use. By video, a teacher can combine students from three schools to form a class and teach them simultaneously even though the schools are twenty miles apart. It's only one of many educational efforts Ameritech has under way. Our Buddy System, for example, is linking students and home computers with their schools and a whole world of information. We're also providing grants to students doing graduate work in telecommunications and supporting teachers writing curricula using our Ameritech PagesPlus® directories as learning tools.

At Ameritech, we're developing innovative products and services of all kinds to help make sure today's students aren't just ready to face the future, but to shape it.

For more information on our contributions to the quality of life in communities we serve, call 1-800/242-8580.

THE COMMUNICATIONS COMPANIES OF AMERITECH ARE:
Illinois Bell • Indiana Bell • Michigan Bell • Ohio Bell • Wisconsin Bell
Ameritech Audiotex Services • Ameritech Credit • Ameritech Development
Ameritech Information Systems • Ameritech International
Ameritech Mobile Communications • Ameritech Publishing
Ameritech Services • The Tigon Corporation.

AMERITECH
Connecting You With Tomorrow

© 1992 SCIENTIFIC AMERICAN, INC

Sighting Cepheid Variables

Not all stars shine as steadily as the sun. A small percentage visibly fluctuate in brightness. But far from being mere curiosities, variable stars offer much information about themselves and the universe as a whole. Among the most useful are the so-called type I Cepheids, because they are good indicators of distance. This property comes about because a definite relation exists between a Cepheid's period of variability and its luminosity: the longer the period, the intrinsically brighter the star. Finding a Cepheid's period and measuring the apparent brightness enable astronomers to deduce its distance from the earth [see "The Expansion Rate and Size of the Universe," by Wendy L. Freedman, page 54].

Getting accurate distance figures is usually not straightforward. Because of the range of magnitudes in the period-luminosity relation, distance calculations can be off by a factor of two or three. To correct for this uncertainty, astronomers generally rely on observations of several Cepheids in proximity. For example, I discovered 29 Cepheids in the galaxy NGC 3109. I fit their apparent magnitudes at maximum light into the graph of the period-luminosity relation [*see bottom illustration on opposite page*]. The fit yielded a distance of 1.9 megaparsecs. Even this figure is only approximate, since it is not corrected for absorption by the intervening interstellar matter. Accurate distance measurements usually demand large telescopes, filters and complicated techniques to extrapolate measurements from one celestial body to another. As such, determining distances is a task best left to the professional.

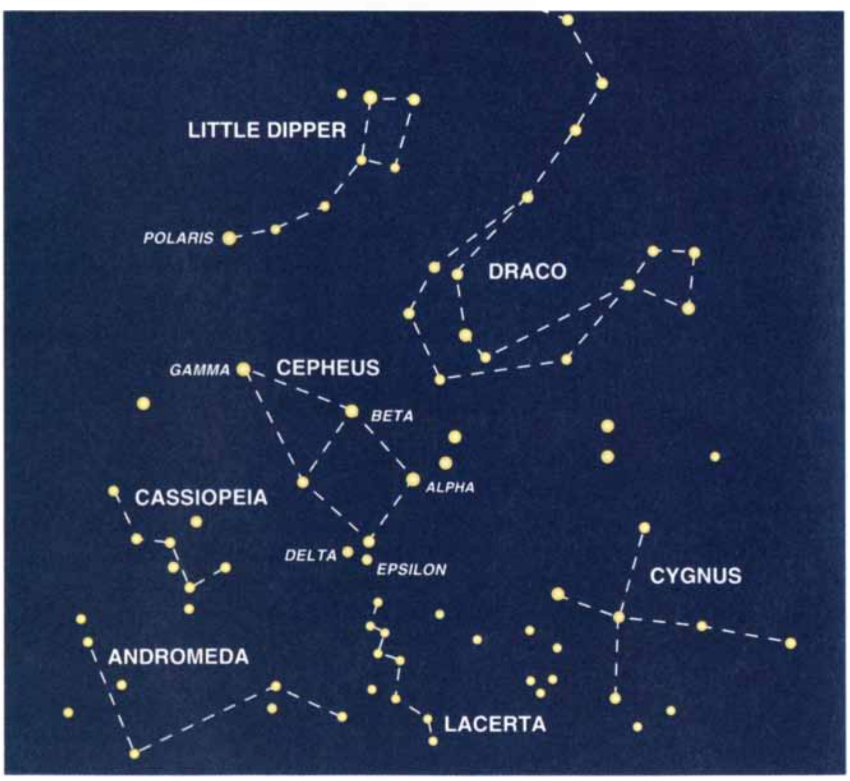
Yet the amateur is not left completely out in the cold. You can readily determine a Cepheid's brightness over time and display the information as a light curve. One of the most useful resources in making observations of variable stars is the *Astronomical Almanac*. This reference (available from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402) lists the variable stars visible to the naked eye. The information includes the names of the stars, their type and location and their period of variability. The almanac also lists a recent time of maximum light, which can be used to predict future maximum brightness.

Stellar brightness is expressed in terms of the magnitude scale. Each unit on the scale corresponds to a luminosity difference of 2.5 times. Fainter stars have larger magnitude numbers; the brightest objects in the sky have negative numbers. For instance, the sun shines at magnitude −26. The unaided eye can detect objects up to a magnitude of about six.

The times given in the almanac are in Julian days, which is a running decimal number related to the dates and times of events. Julian dates are much more convenient than ordinary calendar dates. To obtain the interval between events, you simply subtract the dates. Section B of the almanac lists the equivalent calendar dates.

The general strategy for observing Cepheid variables is to measure the brightness of a Cepheid at many times over many cycles. There are two ways to collect the data on light variation. The first is simply to use the unaided eye and make informed estimates. The second is to photograph the star and determine the brightness by measuring the size of the image.

In either case, begin by making the first set of measurements close together over one or two cycles, so you can get a rough idea of the period. Then wait for a few cycles and take a second set of measurements. But do not wait too long: when the number of cycles multiplied by the uncertainty in the period equals one period, you would be off in the cycle count by one. Pause for several more cycles before taking a third



DELTA CEPHEI is in the constellation Cepheus, near the Little Dipper. Compare its brightness over time with that of Beta and Epsilon Cephei. Astronomers use the Greek alphabet to label stars in a constellation according to brightness.

GEORGE A. CARLSON teaches astronomy, physics, chemistry and piano (music is his other passion) at Citrus College in Glendora, Calif. He is also a visiting astronomer at the Carnegie Institution. He received his Ph.D. in physics from the University of California, Davis.

© 1992 SCIENTIFIC AMERICAN, INC

set of observations. You can repeat this process indefinitely. Knowing the number of cycles between the observations, you can "phase" together your observations to produce an accurate value for the period. I use a computer program to perform the task; the box on the next page describes the algorithm.

The procedure to observe Cepheids with the naked eye is simple in principle and is attributed to the German astronomer Friedrich W. A. Argelander. You select ordinary (nonvariable) stars to compare with the Cepheids. The apparent magnitudes of the comparison stars should cover the full range of the variable's brightness. The comparison stars should be located in the sky as closely as possible to the variable, so that each comparison star and the variable are in the same field of view. Ideally, the comparison stars should differ in steps of 0.3 to 0.5 in magnitude. I label each comparison star in an alphabetical order based on brightness, with "a" being the brightest.

At the time of observation, decide which two comparison stars have magnitudes that bracket those of the variable. Then try to estimate to the nearest tenth the relative brightness of the Cepheid in the bracketed interval. Thus, if you estimate that the magnitude of a variable is 0.4 between those of comparison stars b and c, write b4c as the magnitude (the Cepheid observer's shorthand). You can convert to the actual value by interpolating between the known magnitudes of b and c.

Of course, suitable comparison stars may not lie close enough to the variable. In that case, you would need to shift your line of sight back and forth between the stars to estimate relative brightness. Visual memory is very short; when you concentrate on the second star, your eye has "forgotten" how bright the first one looked. In addition, the twinkling of stars makes it more difficult to judge their brightness. Even so, with practice, your ability to estimate brightness can improve. In fact, with only three levels of brightness, the period of a variable can be found with reasonable accuracy.

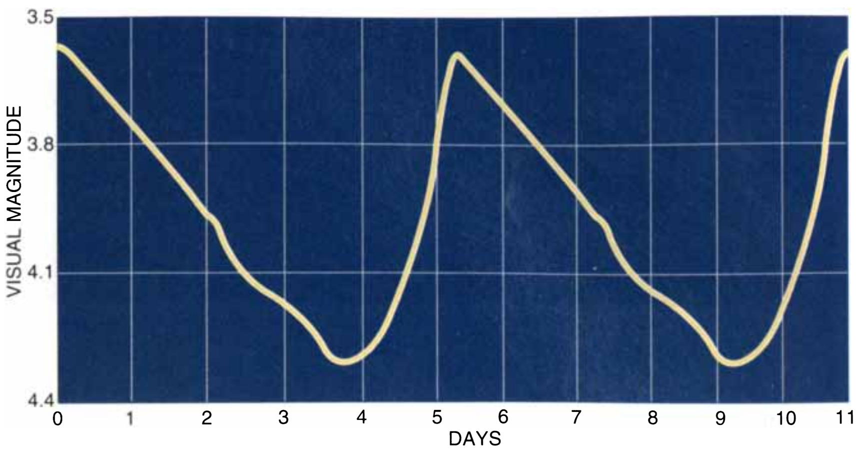
Try the technique on Delta Cephei, which has an apparent visual magnitude of 3.48 at maximum and 4.37 at minimum. Locate Beta Cephei (magnitude 3.23) from a star chart and use it as comparison star a. Use Epsilon Cephei, at a magnitude of 4.19, as comparison star b. Observe Delta Cephei once a night for a week or two, and write down its time and brightness as compared with a and b. If the brightness of Delta Cephei is in between those of a and b, simply write a5b. These three values can provide a crude estimate of the period. With a little practice and some luck with the weather, you should be able to refine your estimates. The almanac lists other Cepheids and suitable comparison stars if Delta Cephei is not easily seen from your location.

A more accurate means to record Cepheid variability is through photography. Because film can detect objects too faint for the naked eye, you can conduct a photographic study of Cepheids not even listed in the almanac. You will need a camera mounted to your telescope and a clock-drive unit (a tracking mechanism that follows the apparent motion of a celestial body).

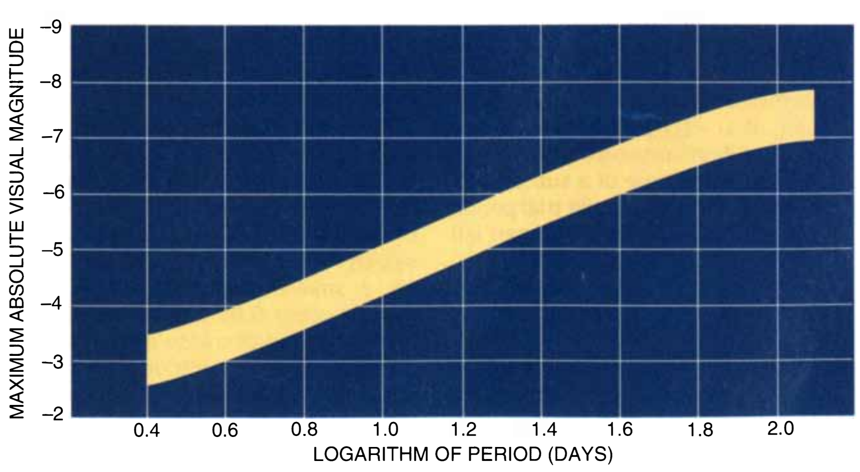
For a variable such as Delta Cephei, a typical exposure through a small telescope would be about 10 seconds. You should, however, take several photographs with different exposure times and select the best time. I recommend doing your own developing so that you can keep the processing consistent from one roll of film to the next.

Film records the light intensity of stars as size: the brighter the star, the larger the image. Thus, you will need to correlate the diameter of a stellar image with a star whose magnitude is known (magnitudes are printed in the almanac and other handbooks). In my astronomy course, I use the open cluster NGC 6940 to teach my students how to perform the task. I use a 12-power comparator (essentially a loupe) to measure the images of the 15 stars in the cluster on an enlarged print. The comparator contains a reticle that has a linear scale with 0.1-millimeter divisions (available from scientific-equipment supply companies). You should be able to estimate sizes to within a few hundredths of a millimeter.

A plot of the magnitudes of these stars against the logarithm of the square of the diameters yields a straight line [*see illustration on next page*]. Once calibrated, you can find the magnitude of

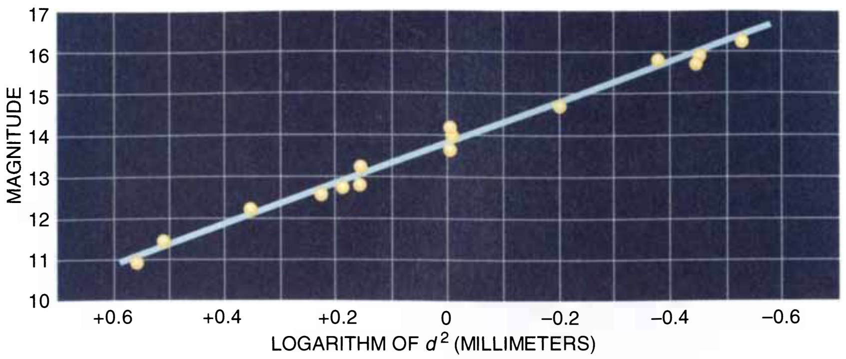


LIGHT CURVE for Delta Cephei shows that the star varies by approximately a magnitude in brightness over a period of about five days.



PERIOD-LUMINOSITY RELATION for type I Cepheids, shown here for visual wavelengths, helps astronomers deduce a Cepheid's distance. Each period has a range of possible magnitudes, which arise because of inherent differences among Cepheids.

© 1992 SCIENTIFIC AMERICAN, INC



***CALIBRATION LINE** for photographed stars in the open cluster NGC 6940 relates the diameters of the images,* **d**, *to the apparent magnitudes.*

any star (including possible Cepheids) in the field of view. Just measure the star's diameter and use the calibration line to convert. Each photograph will have to be calibrated in this way.

Indeed, it may be that the magnitudes of the stars you need for calibration are not listed in any handbook. In that case, you must calibrate those stars against stars whose magnitudes are listed. This secondary calibration needs to be done only once. You can then use these secondary standards against which to measure the variable.

To ensure accuracy in secondary calibrations, you should take certain precautions. First, use the same exposure time for all the photographs. Second, shoot everything on the same roll of film, so that the images experience the same developer concentration and processing time. Third, work from negative prints; it is easier to measure dark images on a light background rather than vice versa. (I make such prints from contact negatives of the original negatives.) Finally, correct for the fact that you are looking along different optical paths through the atmosphere.

The most systematic method is to convert your magnitudes to what they would be if you were looking straight up. For visual wavelengths, the amount of correction is equal to about 0.14 (secant $z - 1$), where z is the angle between the zenith and photographed field. Once you have calibrated your photographs, you can find the period by using the algorithm in the box on this page.

You can construct light curves by plotting magnitude versus phase. Remember to plot the smaller magnitudes (brighter values) toward the top of the ordinate. The conventional way to display light curves is to show two complete cycles.

If you decide to pursue more advanced kinds of observations, you can obtain instructions and software for data analysis from many sources. One of the most useful is the American Association of Variable Star Observers (AAVSO), which can supply finder charts for thousands of variable stars as well as their corresponding comparison stars. The association can recommend projects suitable to each observer's geographic location, equipment, observing conditions and schedule. Interested readers can contact the AAVSO at 21 Birch Street, Cambridge, MA 02138. Amateur astronomy magazines also provide information on Cepheids.

For a sample data set and a copy of the program that finds the period (written in FORTRAN), send a self-addressed, stamped business envelope to the Amateur Scientist, Sighting Cepheid Variables, Scientific American, 415 Madison Avenue, New York, NY 10017-1111.

Finding the Period

I use a program based on one written by Hugo G. Marraco and Juan C. Muzzio of the National University of La Plata in Argentina. The essential features of the algorithm follow.

The first step is simply to take a guess; most type I Cepheids have periods of a few days. Then calculate the phase for each data point—that is, determine where in the cycle you observed the Cepheid at that time. Otherwise, you can easily get the wrong period, especially if you have made only a few observations.

To determine the phase, count the number of complete cycles, N, starting at your earliest observation. Mathematically, N = whole number part of $(t - t_0)/p$, where t_0 is the time of first observation (in Julian days); t is the time of a subsequent observation; and p is the trial period you guessed. The fractional part left over is the phase—that is,

$$\Phi_t = [(t - t_0)/p] - N,$$

where Φ_t is the phase at time t.

Now that you have the phase for each data point, calculate the scattering of the data—in other words, the standard deviation of your data points. Break up the total phase range (0 to 1 cycle) into intervals of, say, a tenth and see how many data points fall in that tenth. For example, you might have three data points whose phase values fall in the interval 0.1 to 0.2. You simply calculate their standard deviation. Perform this task for each interval and then find the average standard deviation for all 10 intervals.

Next you will need to increment the period and repeat the phase and scattering calculations. But be careful: if the increment is too large, you might skip over the correct period without noticing it. On the other hand, if the period increment is too small, it would take a very long time to find the right value.

After covering the period range, look at the values of the scattering and locate the minimum standard deviation. Select a small range of period around this value and search for a more precise period using a smaller period increment. I usually select 0.01 day as my first increment and decrease it by a factor of 10 for each iteration. You can tell when you are finished because the scattering parameter is the same for several successive values of the period. Select the central value as your best period.

FURTHER READING

THE REALM OF THE NEBULAE. Edwin P. Hubble. Yale University Press, 1936 (reprint 1982).

THE NATURE OF VARIABLE STARS. Paul W. Merrill. Macmillan Company, 1938.

CEPHEIDS: THEORY AND OBSERVATIONS. Edited by Barry F. Madore. Proceedings of the IAU Colloquium No. 82. Cambridge University Press, 1985.

THE EXTRAGALACTIC DISTANCE SCALE. Edited by Sidney van den Bergh and Christopher J. Pritchet. Astronomical Society of the Pacific Conference Series, Vol. 4; 1988.

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC

Three-Dimensional Words

BEFORE WRITING, Volume I: **FROM COUNTING TO CUNEIFORM,** by Denise Schmandt-Besserat. University of Texas Press, 1992 ($60).

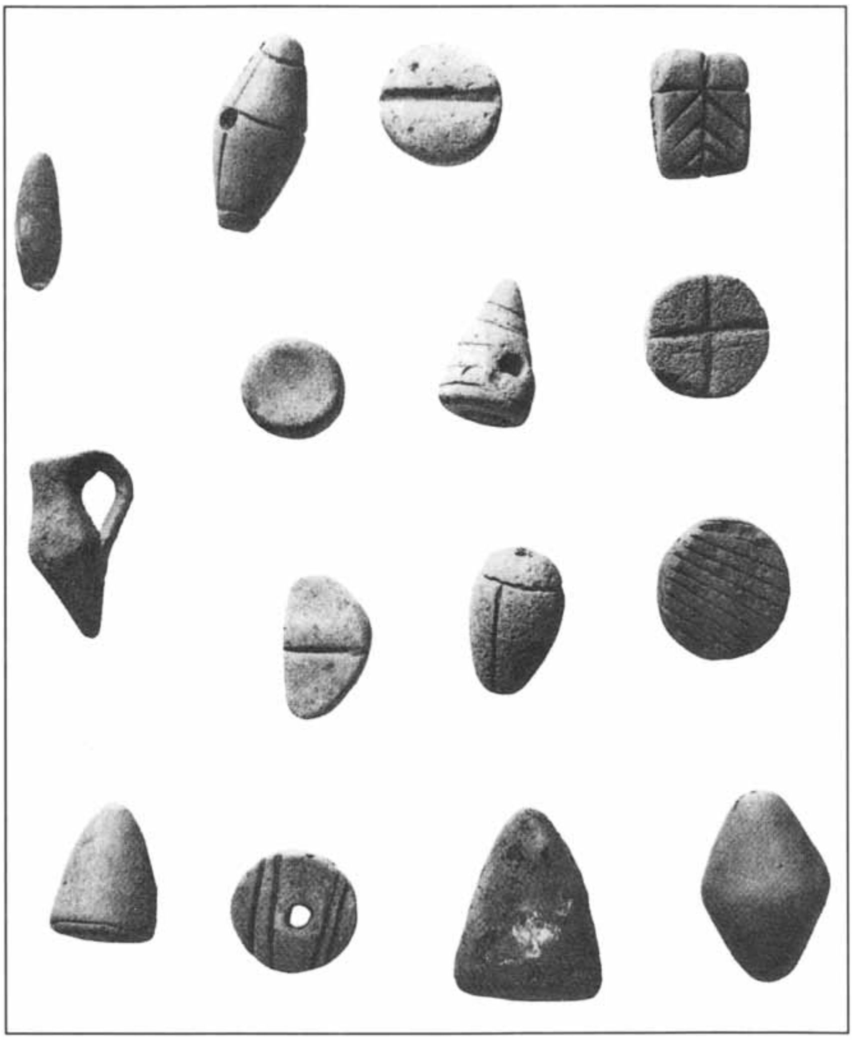
Hundreds of clay tablets marked in cuneiform were found in the 1930s during excavation of Layer III of "the first and foremost Sumerian city," Uruk (the Biblical Erech, in present-day Iraq). The earliest were written in the decades before 3000 B.C. Those archaic texts are surprisingly mature, with few pictographs and many abstract symbols, not much different from texts of Sumer a millennium or more later. The sign for "sheep," for instance, was a circled cross; for "metal," a crescent with five lines. Sophisticated writing had appeared suddenly, ready-made.

The priest-poets of Sumer had their own explanations. In one myth, shining Inanna, the divine sister of mortal Gilgamesh, hero-founder of Uruk, received from Father Enki, God of Wisdom, an imprudently generous gift. "Swaying with drink," he gave her the precious *me,* the 100 elements of all civilization. First on the list was the high priesthood itself! After many indispensable older arts, including lovemaking, song, even treachery, there came eight modern crafts, one among them the craft of the scribe. Inanna, rejoicing, ferried the *me* upriver to her holy shrine at Uruk, where in time the mute clay would be made to speak by well-trained scribes.

This new, scrupulous and exciting archaeological account of how writing came is given us by a brilliant scholar, a woman who of course makes no divine claims at all. What shines here is the human mind, spinning a tight web of inference from abundant evidence. Evidence comes out of 100-odd sites disclosed by a century of the spade, excavations mainly along the Jordan and the three rivers that flow to the head of the Persian Gulf. Not least interesting is the fact that the material she draws on is some 8,000 little hand-modeled pellets of fired clay, often rejected or ignored even during excavation, now become chief cornerstones of her powerful demonstration. (Volume II, not seen or reviewed, is a technical reference catalogue that lists and locates the tokens in their thousands.)

Twenty years ago Professor Schmandt-Besserat was a new postdoctoral fellow at the Radcliffe (now the Bunting) Institute. She wanted to study the uses of clay in the Near East before pottery, bits of clay floors, hearth linings, bricks and more. "Wherever I would go," in museums on four continents, these geometric trinkets were always present. They were the oldest clay objects to have been fire-hardened. The very earliest ones may have been "baked in domestic ovens," but the latest show perfectly controlled and much hotter firing. Most archaeologists had ignored them; a few had jumped to the unsupported conclusion that they were amulets or game pieces.

She came to call them tokens. A Rosetta stone for the tokens had been found by 1960, but not fully read. A peculiar empty "hollow tablet" bore a late cuneiform inscription that described once enclosed counters that listed so many ewes, lambs and sheep. The count in the textual list matched the number of "stones" reported to have been held in the hollow tablet when the excavators had opened it. But the counters themselves were neither saved nor described. A second hollow tablet was reported in 1966 by the author's teacher, Pierre Amiet, at the Louvre. From Susa, it was much older, preliterate, and it still held its original contents, the very sort of geometric artifacts Schmandt-Besserat would later find in abundance lying loose and out of context on museum shelves. "In 1970, two pieces of the puzzle snapped together for me...I had not



TOKENS from Uruk in southern Mesopotamia (present-day Iraq), 8000 to 3500 B.C.

© 1992 SCIENTIFIC AMERICAN, INC

seen [the Amiet article] since I began collecting tokens. I could not believe my eyes when I saw the small clay cones, spheres, and tetrahedrons" in Amiet's illustrations. The use of that multitude of tokens had been found.

Tokens were a common code. They are ideal symbols, save only for their three-dimensional nature. They are discrete, recognizable, repeatable, durable, cheap, yet open-ended enough to allow many new forms. They were a record-keeping device at village scale, one that "swept across the Near East on the coattails of agriculture" for 5,000 years, remarkably free of any regional variations until cities began. At first there were at most a dozen or two simple forms. Then they entered a second, more complex phase, to dwindle once writing had come.

The first two hollow tablets reported have been followed by 115 more, most later than 4000 B.C. Now these are recognized as envelopes of clay. Nearly all the envelopes are covered with repeated seal impressions, a signature—sometimes several—authenticating the security of the contents. Only a few have been opened to check the contents (x-ray techniques have not yet given good results). The number of tokens within is never very high; on the average there are about nine. These are no records of large-scale trade but rather of villagers' contributions to pooled grain or livestock surpluses, subject to some later redistribution. Step by step, such communities became "ranked societies," in which redistribution allows in the end for a tribute of offerings, fees and taxes.

Next come well-marked envelopes. For them, many tokens have been impressed into the outer clay, to signal to the scribes just what is in the authentic record sealed within. The match in significance between token types and farm products (rather than, say, days of labor) rests on the subsequent cuneiform symbols. These can often be traced back in form to impressions made on clay tablets a few centuries before stylus-written cuneiform proper appears. Those early impressed signs match nicely the old forms of the tokens and often enough fit the epigrapher's judgment of the origin of the later, more stylized cuneiform whose meaning is known from rich literary context.

After 4000 B.C., "industry gave a major boost to the token system." Hundreds of new types are found, largely in Uruk. Some of them are quite figurative, tiny bowls, jars with handles, even little trussed ducks, all of them handmade clay pellets, easy caricatures, often incised with simple dots and stripes. (A few are even molded.) These changes extend the token code to an urbanized society. The temple bureaucrats were no longer content with the simple tributes of the farmer but drew their levies from a variety of urban craftsmen.

Quite another dimension lies within this chronicle of the well-known rise of city-state, class and the division of labor. It is the growing abstraction and complexity of the code. Perhaps the first recorded counts came in the upper Paleolithic, simple tallies of time passed, say, one mark on a bone for each day of the lunar month. Later came tokens, every token form a three-dimensional noun, each commodity counted concretely by its set of identical tokens. Nature and quantity were still fused.

The new step is found on inscribed tablets of Sumer during the last centuries before writing: abstract count. Number was no longer embodied by a single form class of tokens but stood in two dimensions as a set of marks on clay. The tablets bore impressed numerals, at first simple tallies like the Roman numeral III. Adjoining each such impressed numeral lies an incised mark, often the drawing of a complex token form, to identify what was counted. Arithmetic had come, not merely a count of days or of ducks but pure number itself, that class of classes.

The accountants invented the first numerals on clay around 3100 B.C., encoding the concepts of "oneness, twoness, threeness." In the city stage called Uruk VI, it took one three-dimensional ovoid token to record one customary jar of oil. A little later, in the overlying layer called Uruk IVa, it took two markings on the surface of a clay tablet. One impression recorded a customary unit measure, the jarful, using the outline form of the old ovoid token. A second single, strong mark conveyed the pure numeral 1: one oil jar. By Uruk III, it took three signs, one for the numeral 1, one for the standard jarful and a new symbol that denoted oil itself. With three signs, a flexible written language had arrived.

Once individual citizens needed to be identified, the idea of names in rebus was not far off. Given one mark or placement convention, like the cartouche in Egyptian hieroglyphics, a string of symbols could stand no longer only for concept or commodity but simply for the sound connoted through the spoken language: phonetics, that novelty, almost the last decisive one, depended less on the social milieu than on internal practice among specialist scribes. Within a few centuries they had invented the way we now share the myth of Inanna and the sustained argument in this absorbing book.

In Praise of Pauling

THE CHEMICAL BOND: STRUCTURE AND DYNAMICS, edited by Ahmed Zewail. Academic Press, 1992 ($49.95).

Certainly chemistry is structure: the double helix and the hexagonal benzene ring have become logos for the entire science. But chemistry unfolds in time as it dwells in space. Reactions, the *change* of chemical structures, are equally at the roots of the science. And there is a final necessity: those clever chemists themselves. This fine book introduces all three.

Its authors are nine celebrated chemists from North America and the United Kingdom. Six of them are Nobel laureates; one, Linus Pauling, is a rarity indeed, a Double Nobelist, once for chemistry, once for peace. The attractive volume grew from a splendid occasion, the symposium in February of 1991 when Caltech celebrated Pauling's 90th birthday. Personal reminiscences, colorful diagrams and photographs and biographies of the contributors welcome the general scientific reader to nine relatively informal chapters.

Linus Pauling opens by recalling his early interest in minerals at age 12, when in wonder he collected the local agates. At 18 he became an assistant instructor in quantitative chemical analysis at Oregon Agricultural College, to give one of the two seminars that year. One man spoke on the frozen fish industry; young Pauling told of the electron theory of the chemical bond from the recent papers of G. N. Lewis and Irving Langmuir that had caught him. A little later he tried to deposit single crystals of iron in a magnetic field: "without success." But it was enough to spark expression of an interest in crystallography, and Professor Arthur Amos Noyes of Caltech proposed to the just-accepted teaching fellow that he begin graduate research there with x-ray diffraction.

For a dozen years Pauling made and analyzed Laue photographs and thought deeply about structures. "In 1934 the transition...to modern x-ray crystallography was begun by the...Patterson diagram." That boldly simplifying approximation to the location of atoms in lattices was just what so good a chemist could use. Pauling also mastered the old quantum theory, found some of its limitations, and made his way as a fresh Ph.D. to a fellowship in Munich with Sommerfeld in the spring of 1926, just when Schrödinger published the first papers on the wave equation. The young American returned a working quantum chemist, whose steady flow of insights and approximations has become classic.

© 1992 SCIENTIFIC AMERICAN, INC

"When the earthquake hit, I watched helplessly as my poor LaserJet went crashing to the floor."

—James Cost, Chief of Police, Campbell, CA

"The LaserJet fell from about four feet," said Police Chief James Cost, recalling the disastrous San Francisco earthquake. "I assumed we'd have to buy another one. But later, my secretary plugged it in and printed a document. It's about the only thing that did work that day."

Reliability is only one of many reasons more people use HP LaserJet printers than all other laser printers combined. Some people cite innovations like Resolution Enhancement technology, which creates sharper edges and better overall print quality. Or genuine PCL5 printer language. Or the range of options, with five printers priced from $1,249 to $5,495.*

To hear what other customers have to say about HP LaserJet printers, call **1-800-752-0900, Ext. 2987** for a free video. And see why so many people have an unshakable faith in HP quality.

HP LaserJet Printers.

HEWLETT PACKARD

*Suggested U.S. list prices.
©1992 Hewlett-Packard Company PE12253

© 1992 SCIENTIFIC AMERICAN, INC

In 1939 Pauling published a famous series of lectures on the chemical bond, the deepest view we then had. Nobelist Max F. Perutz read chemistry then in Vienna. That meant he wrote down what the professor said, "or rather my girl friend who knew shorthand did, and I learnt it by heart." Pauling's book was not at all like that. It included his opinion that the easily formed, rather weak hydrogen bond, one proton lying between two somewhat negative atoms, would be more significant for the molecules of life than "any other single structural feature." Postwar, Pauling would add that hemoglobin might be the molecule whose mapping could tease out the full nature of protein structure, in spite of the prodigious amount of work. Perutz read it all and says, "I took this admonition to heart, but it took me another 30 years to do the job."

Francis Crick read the same book: "It is almost true to say that's the only chemistry I ever learned." Pauling's insistence on the basic importance of quantum chemistry was decisive for Crick. Physicist Max Delbrück, pioneer of molecular biology, thought, like Niels Bohr, that there would have to be new physics within the giant molecules. When he saw the double helix of DNA, "he thought it was too much like a tinkertoy." But Pauling expected that directness of fit. He had a long string of profound structural ideas: the alpha helix, pleated sheets, enzyme sites and coiled coils. But also he led an active group who worked hard at real and meaningful structures, the foundation of molecular biology.

The topic of another Pauling admirer, Alexander Rich, is the double helix in three forms. First is the DNA we know, then the same DNA without solvent water and, last, the left-handed, less stable, Z-DNA. Grooves in the three molecules offer clues to how proteins recognize their instructive sites.

The chemists were slow to split seconds. In 1947 the free radicals, transient intermediate steps in most reactions, seemed beyond direct chemical study. Microsecond chemistry was begun with a flash by another author, Lord George Porter, in 1949. Start a gas reaction with a brief flash, and probe the changes without delay by a second tailored flash, analyzed by apt spectrometry.

Step by step, that technique has been extended down nine orders of magnitude, most recently by editor Zewail, the Linus Pauling Professor of Chemical Physics, at work in Pauling's old x-ray lab at Caltech. A photograph of his tamed beams zigzagging around a big floating optical table is included. The ingenious apparatus can record chemical identifications within tens of femtoseconds by use of ultrashort laser pulses, collisionless, polarized molecular beams, and well-timed pulse manipulation. Such studies show directly "the ephemeral, but all-important, transition states in chemical reactions" and spatially resolve atomic interactions down to a tenth or less of atomic dimensions. In his own chapter, John C. Polanyi, the modern inventor of the transition state, elaborates the subtle idea helpfully, and another chapter outlines the complex multidimensional energy surfaces many reactions imply. (Here and in a few other places the up-to-date material makes demands most nonchemist readers will not easily meet.)

Elder Pauling does not avoid the fray. He cites new experiments and theory to support his view that the intermetallic quasicrystals of the experts, with their unendingly frustrated repetitions, are in fact much simpler twinned structures that only simulate the fivefold symmetry denied all true crystals. Acute, feisty, as astonishing as ever, hero-chemist Linus Pauling may—or may not—be right once again, at a ripe 91.

Big Bangs

Why Buildings Fall Down: How Structures Fail, by Matthys Levy and Mario Salvadori. Illustrations by Kevin Woest. W. W. Norton and Company, 1992 ($24.95).

In this insiders' book, we are let in on a lot of delicious trade talk that frames the specific explanations. Using drawings in plenty, but not even one drop of algebra, the account brings you to see why buildings stand up and "yes, but once in a blue moon, [why] they fall down." These are two savvy old pros, distinguished structural architects and principals in a New York consulting firm, noted for detective skill in investigating buildings that failed under certain newsworthy blue moons. Professor Salvadori years ago established his talent and verve as an expositor in a memorable book or two on why buildings stand up; this new volume has as much insight and moves with the swift action of the Johnstown Flood (not ignored) across dozens of celebrated examples.

Theirs is no new problem. Of the Seven Wonders of the Ancient World, what with wars and earthquakes and vandals and neglect, six have fallen. Only the Great Pyramid still stands. One early pyramid did in fact slough off all its two-ton limestone masonry casing blocks; its foundation design explains the failure, not repeated. Poor King Kong of course

© 1992 SCIENTIFIC AMERICAN, INC

did not overstress the Empire State Building—his tragedy was only in special effects—but that tall, redundant steel frame literally shook off the real impact of a fog-lost bomber in 1945. Head-on collision at speed just above the 79th floorline killed the crew and 10 people in the building, mostly by burning gasoline. (The feisty Little Flower came up behind the fire fighters and was seen shaking his mayoral fist "and muttering: 'I told them not to fly over the city.' ")

The massive tower shook mildly under the jolt—the 10-ton bomber brought in only half a percent of the momentum of the design wind load—and after one double sway, witnesses recalled, the motions settled down. Masonry-clad towers are well damped by friction between the steel elements and the heavy walls, but lighter modern towers vibrate much more easily. The beautiful and now truly safe Hancock Tower, whose mirrored sunsets so vivify the Boston skyline, in 1973 began a perilous window-shedding motion in response to local windstorms. A cross fire of litigation, all against all, ended in a legal agreement of "nondisclosure in perpetuity," but the fraternity of builders, tight-lipped to outsiders, is a "first-rate grapevine to its members."

The tale leaked here is not all new, but it is particularly well told. Everything went a little wrong. One major task was to stabilize the unusually shaped tower against its twisting response to wind. The strong glass windows cracked as the frame swayed; more than 10,000 of them were replaced with subtly less vulnerable ones. More steel alone would not do; the vibrations had to be damped as well. The 58th floor of the tower was given over to two 300-ton blocks of lead, each free to slide on an oily platform. The building in effect slips under the massive blocks, whose inertia keeps them from quick response. Big springs, tuned to precisely the right rhythms, transfer the motions of the tower walls to the two blocks; the motional energy the blocks gain is steadily drained out to become mere heat within big shock absorbers. The building now sways only a little as the blocks within swing to and fro through many feet.

The senior author's recounting of the tough Q and apt A during two of his appearances on the witness stand is engaging. He has come to admire the wisdom of "uneducated" juries more than he smarts at the barbs of the trial lawyers. A real expert like Mario Salvadori knows that "what counts in a court debate is the whole man rather than the specialist." This appealing volume closes with a 40-page appendix that offers an admirable primer of structural theory.

 by Michael Schulhof

Why Business Needs Scientists

Twenty years ago I was a physicist working on neutron-scattering experiments at Brookhaven National Laboratory. Now, as the vice-chairman of Sony USA and president of Sony Software, I represent Sony in both the electronics and the entertainment business. I spend my days discussing and overseeing projects that range from new developments in high definition to the cutting edge of popular music.

My experience has convinced me that a background in pure science is an ideal preparation for business. I will take that a step further and say that American business would be a lot better off if it had more scientists and fewer M.B.A.'s running its corporations.

Why do I think the neutron detector prepared me for life at Sony? As a physicist, I was doing work I considered important and working with people I admired. But as I looked around the lab, I asked myself whether this was what I wanted to be doing 20 years into the future. I thought I might like to try business, but I was not absolutely sure. When I shared my uncertainty with my thesis adviser, the distinguished researcher Robert Nathans, he gave me some advice I will never forget. "Don't worry about it, Mickey," he said. "You're a physicist. Physicists don't do anything they really don't want to do. If you get into business and find you don't like it, you'll get out."

Obviously, I liked it. I stayed. But I stayed as a physicist. No matter what it says in my job description, I am still a scientist. And I have approached business problems the same way I approached scientific problems. The lessons I learned as a scientist were excellent instruction for business.

Some of those lessons are as basic as a strong work ethic. The business school yuppies of the 1980s glamorized the idea of working long hours. But that trend was in fashion in labs long before anyone ever heard of Michael Milken. I can well remember sitting up until 3 A.M. baby-sitting our precious high-flux beam reactor through an experiment. The hours didn't matter. It was the result that counted. When you have a meaningful challenge, personal time means very little. That is a lesson I have carried over into corporate life.

Science also encouraged my intellectual curiosity. Of course, that was something that attracted me to physics in the first place. But working in the lab at Brookhaven taught me how stimulating it was to make intellectual curiosity the center of your professional life. My responsibilities have obviously changed. But intellectual curiosity is very much a part of what keeps me going in the business world. In science, you accept intellectual curiosity as a given. I wish it were more common in business.

I would also like to see business people develop some of the tenacity that is common in science. People in business tend to be impatient. The scientists I worked with were anxious to see results. But they realized that you had to build the foundation before you could put on the roof. By example, they taught me the importance of mastering the fundamentals of a field before you could do meaningful new work. Shortly after Sony acquired Columbia Pictures, I began to read the scripts for films we had under production. That didn't endear me to some of the operating people. One of them challenged me about why I wanted the scripts. He as much as told me that they were not going to let me take over the creative decisions. But I told him he was missing the point. I was not interested in telling the creative experts how to make films, but I was intensely interested in understanding the process.

Learning as much as you can about the details is a lesson that is actually discouraged in many business schools. They promote the misleading idea of the generic manager—the consummate professional whose education has prepared him or her to step into any kind of business and run it.

The myth of the plug-in executive created a generation of migratory managers in American business. Most of them do not have the time or the inclination to learn anything in-depth about the business they are responsible for. Instead they bring their business school theories to each assignment. And quite often they do not stay around long enough even to evaluate whether or not the theories are valid. That is a big difference between business graduates and science graduates. The business graduates accept theory as gospel. The science graduates accept theory as the starting point for experimentation.

An equally dangerous trend in the graduate schools of business is their potential to restrict creativity. And the greater the reputation of the business school, the greater the risk that its graduates will rely on management theory instead of personal creativity. There is a time for doing things the Wharton way or the Harvard way. But there is also a time for doing things your way.

To be truly successful in business, you have to be a creative risk-taker. I have spent about $7 billion of Sony's money to acquire companies such as Columbia Pictures and CBS Records. These were strategic acquisitions that supported our long-term vision for Sony. You have to have your own vision of the future. And you need the confidence to invest in that vision. It is not much different from the approach to scientific research. The people I admired most in science had the creativity to develop long-term visions of the future as well as the courage to stick with that vision unless research proved them wrong.

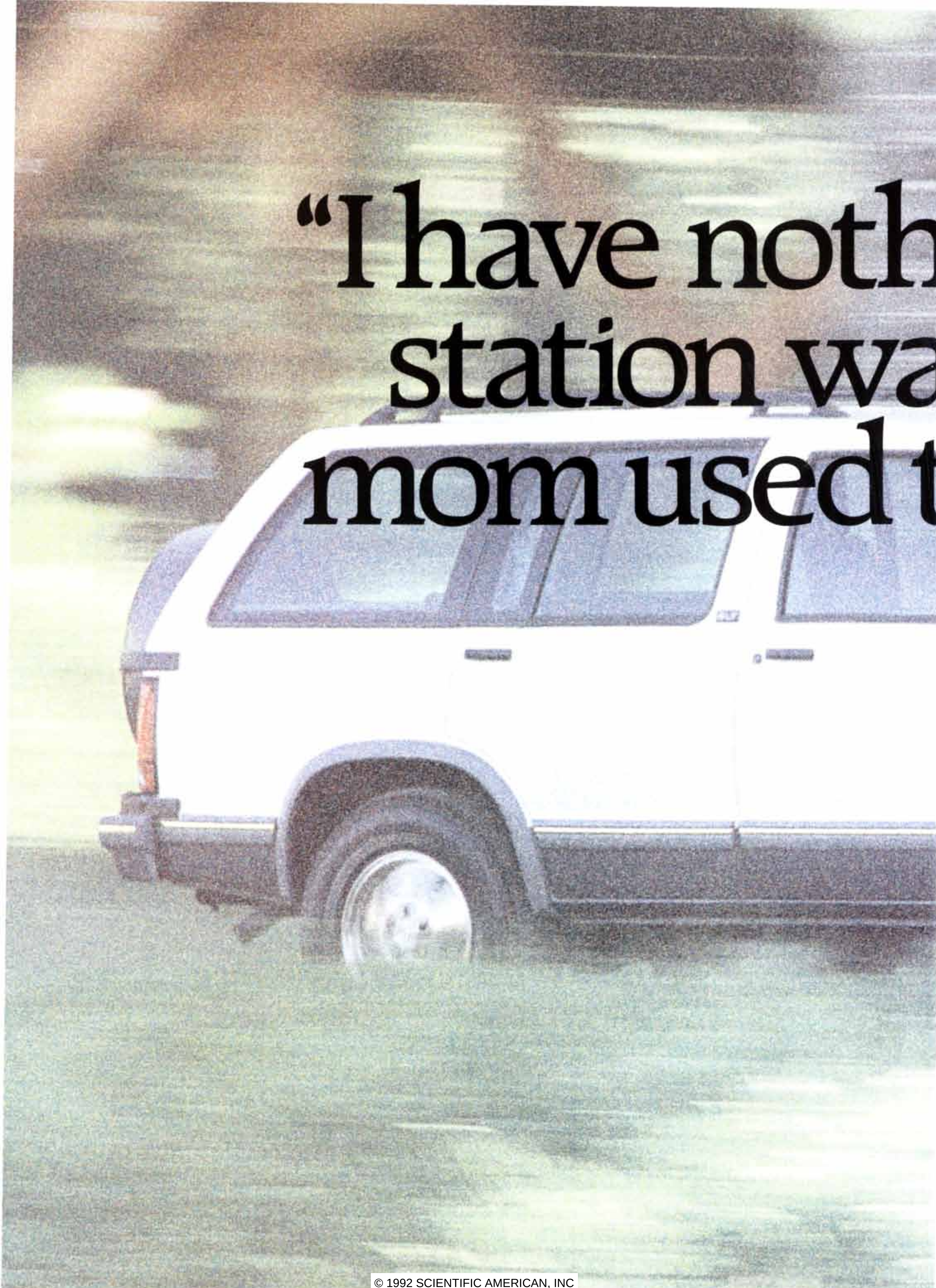
In the years ahead, business people will be asked to solve complex problems with very high stakes, not just for their corporations but for society as a whole. Some of those problems will involve decisions about technology, about the environment, about the economy and the marketplace, even about government. Scientists understand the process of critical thinking. They know how to analyze problems by concentrating on the important elements and filtering out the irrelevant. They understand that worthwhile results require a long-lived effort. They are willing to admit there are things they do not understand and then take the time to find out what it is they don't know.

Business needs that kind of vision and that kind of intellectual courage. Business could get that kind of thinking by taking some of its surplus M.B.A.'s and sending them back to school for Ph.D.'s in science. Fascinating, but unlikely. Instead I think business has the responsibility to recruit more scientists.

MICHAEL SCHULHOF is the vice-chairman of Sony USA and president of Sony Software.

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC



© 1992 SCIENTIFIC AMERICAN, INC

The sedate practicality of a station wagon is meant for someone, but that someone isn't you.

For you there's a more expansive practicality. In a vehicle that seats you comfortably above the hosts of sedan drivers and, with the push of a button, shifts into 4WD to release you from the bounds of the road. A trim, agile vehicle, in which truck power and strength, leather-trimmed seating, 4-wheel anti-lock brakes and advanced Softride suspension are seamlessly blended, presenting a confident view of the world that can only be brought to you by a truck company.

It's called the Jimmy SLT, from GMC Truck. For more details, give us a call at 1-800-GMC-TRUCK (1-800-462-8782). It will not only expand your options in life. But your horizons as well.

GMC TRUCK®
THE STRENGTH OF EXPERIENCE

© 1992 SCIENTIFIC AMERICAN, INC

© 1992 SCIENTIFIC AMERICAN, INC