SCIENTIFIC AMERICAN

HUMAN-POWERED HYDROFOIL

$2.50

December 1986


performance sedan
ontiac can't match.
$9852.
acts for anyone in search of a serious performance sedan.
ith Lancer you can count on a year or 50,000 mile Protection an.** The Honda and Pontiac on't even come close.
Lower price.
bviously, your bottom line consid- ration is price. At just $9852,† Lancer one four-door sedan that delivers e goods for a lower base price an either of the four-door sedans
we've discussed. Lower than the Honda. Lower than the Pontiac.†
Considering all the facts, we think you'll agree that it pays to see your Dodge Dealer and buy or lease the one sedan that takes your bud- get as seriously as your need for performance. Dodge Lancer.
Dodge 5/50
DIVISION OF CHRYSLER MOTORS
Setting new standards of performance.
cks (combined) designed & built in No. America and sold Oct.-Nov. '85 (5 mos. avg. usage) and on a warranty comparison of competitively priced models. See
uminum wheels shown are $310 extra.
BUCKLE UP FOR SAFETY.


Happy Holidays
ABSOLUT ANTICIPATION.
FOR GIFT DELIVERY ANYWHERE CALL 1-800-CHEER-UP (EXCEPT WHERE PROHIBITED BY LAW)
80 AND 100 PROOF/100% GRAIN NEUTRAL SPIRITS (ABSOLUT COUNTRY OF SWEDEN™) © 1986 CARILLON IMPORTERS LTD., TEANECK, NJ

Established 1845 **SCIENTIFIC AMERICAN** December 1986 Volume 255 Number 6

ARTICLES

38 **CORPORATE MEDICINE FOR PROFIT, by Donald W. Light**
The new hospital chains thrive on the health-care system's inefficiencies and highlight its failings.

46 **CHAOS, by James P. Crutchfield, J. Doyne Farmer, Norman H. Packard and Robert S. Shaw** Yes, there is order in chaos. Even random events have an underlying pattern.

58 **DARK MATTER IN THE UNIVERSE, by Lawrence M. Krauss**
The universe contains many times more matter than telescopes can see directly. What is it made of?

88 **THE FIRST HUMAN RETROVIRUS, by Robert C. Gallo**
Novel techniques led to the author's discovery of the virus family that includes the agent of AIDS.

102 **THE FUNCTIONAL ARCHITECTURE OF THE RETINA, by Richard H. Masland**
This remarkably compact bit of neural tissue mobilizes 50 kinds of cells to encode the visual world.

112 **HOW ROOTS RESPOND TO GRAVITY, by Michael L. Evans, Randy Moore and Karl-Heinz Hasenstein** How does a plant root distinguish up from down on the earth but not in outer space?

120 **HUMAN-POWERED WATERCRAFT, by Alec N. Brooks, Allan V. Abbott and David Gordon Wilson** Not racing shells but odd craft that look like winged bicycles are setting records now.

132 **A ROMAN APARTMENT COMPLEX, by Donald J. Watts and Carol Martin Watts**
A rigorous geometry that was rich in philosophical significance underlay every aspect of its design.

DEPARTMENTS

4 LETTERS
10 50 AND 100 YEARS AGO
12 THE AUTHORS
14 COMPUTER RECREATIONS
22 BOOKS
72 SCIENCE AND THE CITIZEN
140 THE AMATEUR SCIENTIST
148 ANNUAL INDEX
152 BIBLIOGRAPHY

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. 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. SUBSCRIPTION RATE: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017

THE COVER

The painting on the cover shows the human-powered watercraft *Flying Fish II* at racing speed, when it is no longer supported by its two pontoons but by its two hydrofoils, or underwater wings. The craft is an example of the new designs that are challenging the traditional shells, or racing rowboats, for speed records on water (see "Human-powered Watercraft," by Alec N. Brooks, Allan V. Abbott and David Gordon Wilson, page 120). A rider on *Flying Fish II* can cover a standard 2,000-meter course about 10 seconds faster than the record set by a single rower in a shell. Its top speed over shorter distances is about 13 knots, or 6.5 meters per second. The craft is ridden much like a bicycle, and a bicyclelike drive chain running through the hollow main-wing strut connects the pedals to a high-efficiency propeller. The handlebars are connected to the smaller front-wing strut, which also serves as a rudder. A spatulalike device that skates along the water behind this strut controls the depth of "flight" by adjusting a flap on the wing.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source	Page	Source
15	David Di Francesco, Pixar	91	Bernhard Kramarsky
16	Lucasfilm Ltd. (*top*), Johnny Johnson (*bottom*)	92–95	George V. Kelvin
		96	Fernando De Noronha
18	David Di Francesco, Pixar (*top*); MacDonald-Dettwiler and Pixar (*bottom*)	97–98	George V. Kelvin
		103	Richard H. Masland
		104	Julie H. Sandell
20	David Di Francesco, Pixar	105–106	Alan D. Iselin
39	National Medical Enterprises, Inc.	107	Richard H. Masland (*middle and bottom*)
40	Jerome Kuhl	108	Richard H. Masland
41	New York Academy of Medicine	109–110	Alan D. Iselin
		111	Ramon F. Dacheux and Elio Raviola
42–45	Jerome Kuhl		
47	Bill Sanderson and James P. Crutchfield	113	Randy Moore
		114	Patricia J. Wynne
48–49	Andrew Christie	115	Randy Moore
50	James P. Crutchfield	116–118	Patricia J. Wynne
51	Andrew Christie (*top*), James P. Crutchfield (*bottom*)	121	Martyn Cowley
		122–126	Hank Iken
52–53	James P. Crutchfield	133	John Moss, courtesy of the Archaeological Superintendency of Ostia
54	Harry L. Swinney and Anke Brandstäter, University of Texas at Austin	134–138	Tom Prentiss
		139	Tom Prentiss (*top*); John Moss, courtesy of the Archaeological Superintendency of Ostia (*bottom*)
55–57	James P. Crutchfield		
59	Palomar Observatory		
60–61	George Retseck		
63	Michael J. Kurtz		
64–68	George Retseck	141–147	Michael Goodman
89–90	George V. Kelvin	144–145	© Minotaur Designs

LETTERS

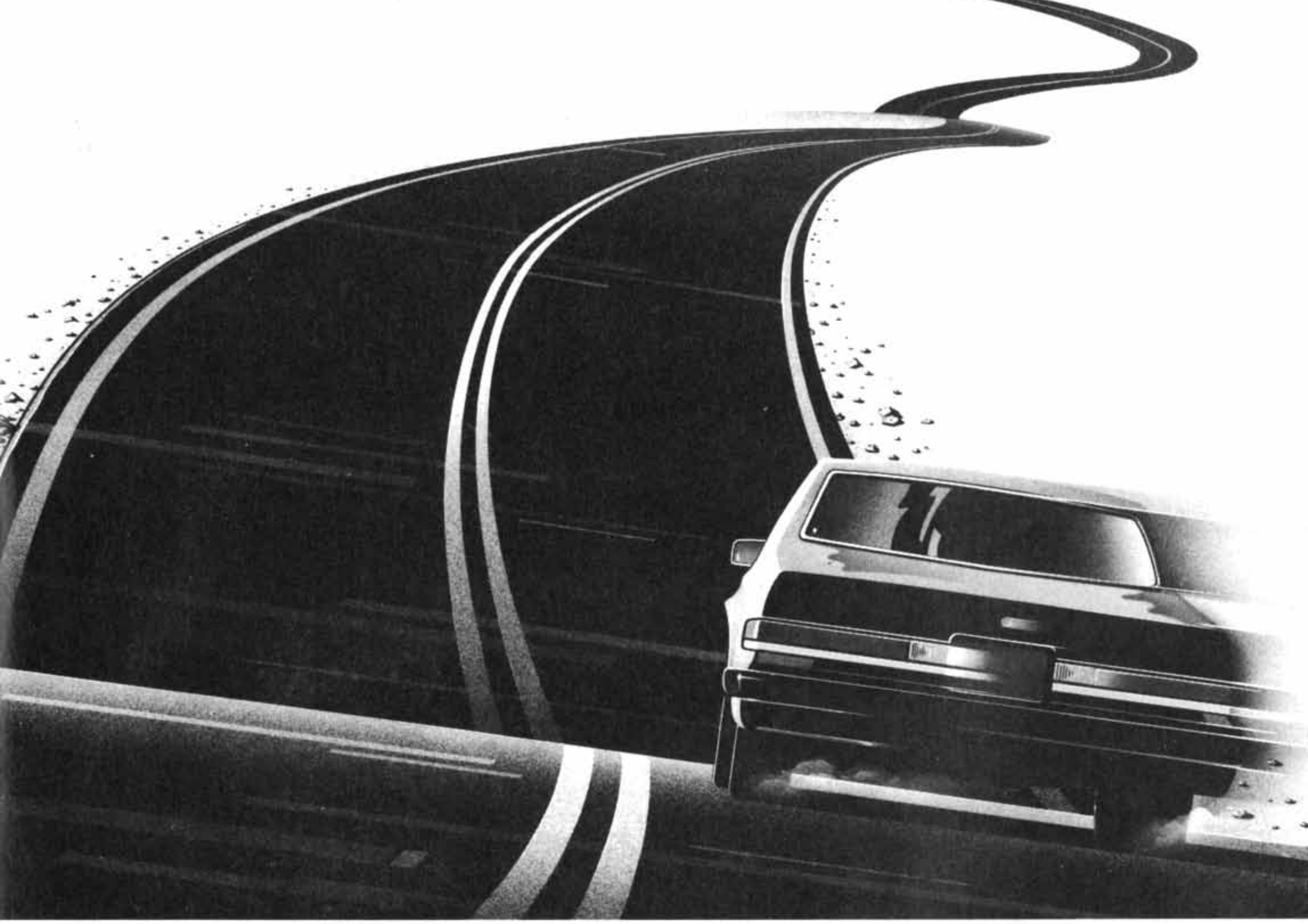
To the Editors:

In "The Microwave Problem" [SCIENTIFIC AMERICAN, September] Kenneth R. Foster and Arthur W. Guy discuss the question of the health hazards of microwaves, give their assessment of the biological literature and state there is no convincing evidence of hazard. But they do not reveal a number of facts the reader should know in order to assess the information the authors provide.

They should have revealed that one of the authors has for years been testifying in legal proceedings on the health hazards of microwaves on behalf of various manufacturers and users of microwave equipment. To cite a few examples, he testified, or was scheduled to testify, in the matter of *Massachusetts Department of Public Health* v. *RKO General, Delsesto and Spinella* v. *Amana Company, Schuerman* v. *Pacific Telephone and Telegraph Company* and *Standiford Zoning Case 85–96.* The authors specifically refer to the case involving the health concerns of the residents of Vernon, N.J., and imply that the residents' fears were unfounded. They do not say that one of the authors was retained by a major microwave-equipment manufacturer and testified on its behalf in that case. Thus these authors are not disinterested parties providing information on health hazards of microwaves.

Moreover, the authors base much of their article on the ANSI (American National Standards Institute) "standard." They do not tell the reader that ANSI is almost entirely funded by industry. Its officers and directors are almost entirely drawn from industry. The military, the biggest user of microwave equipment, initiated the ANSI microwave-biohazard committee, selected its first chairman and has largely determined what research would be carried out on the biological effects of microwaves. The authors do not reveal that conflict of interest or the fact that most of their own research funds have come from the military.

The authors misrepresent the findings and status of biological research on the effects of low-intensity microwave energy. As a biological scientist, part of whose basic research has included the use of microwave energy, I can recognize a rather clear and consistent pattern in the data on biological interactions with low-intensity microwaves. I detail some of the findings in "The Evolution and More Significant Results of Biological Research with



All you asked was to go faster and farther. With more power. And less fuel.

In many of the finest American cars, Bosch fuel injection components help deliver precisely the amount of fuel needed for any driving condition.

And those American cars are going faster and farther. Producing up to 20 percent more horsepower. Using less fuel. And reducing exhaust emissions.

Power as you need it.

Bosch multi-point systems inject fuel right at the individual cylinders.

A microcomputer reads thousands of engine operations and driving conditions every second, and dictates the most efficient mix.

So Bosch fuel injection not only gives you more power when you need it. It can boost your mileage by as much as 15 percent.

Traveling in the best of companies.

Bosch fuel injection arrived in the U.S. in imported cars over 30 years ago. It's since evolved from a curiosity to a high-technology must for today's best performing machines.

Today, every U.S. manufacturer uses Bosch fuel injection systems or components.

Virtually every European model sold in the U.S. is equipped with Bosch fuel injection.

And every fuel-injected Japanese model uses systems or components produced under Bosch license or using Bosch patents.

Bosch breaks ground.

Sophisticated problem solving is what Bosch is all about.

We supply every major U.S. and European automaker with an array of advanced products, from high-performance spark plugs to anti-lock braking systems. But our expertise isn't limited to automotive.

We're a leading manufacturer of high technology products, such as Bosch factory automation systems, packaging machinery, medical equipment, home appliances, power tools, TV studio and video graphics equipment, and Blaupunkt automotive sound systems.

In fact, we've pioneered literally thousands of the world's most sophisticated electronic and electromechanical advances. As a result, we're an international corporation with manufacturing facilities in 18 countries. And with sales and service in 139.

But more important are 147,000 Bosch people—9,000 in R&D alone—who are capable of finding the right answer to the toughest challenges. Including yours.

BOSCH

Bringing high technology down to earth.

Low Intensity Non-ionizing Radiation," a chapter in the forthcoming *Handbook of Bioelectricity*. There are also numerous published papers that show clear-cut interactions of low-intensity microwave energy with organisms and represent significant advances in basic biology. But the authors do not discuss this research, for they are neither disinterested parties nor biologists.

Finally, the authors refer to ambiguity in the implications of the biological data for human health. In this brief letter there is no space to cite all the misinformation in their account of the research on cancer, the blood-brain barrier, hearing and the heart. Nor is there space to cite the misinformation in what they say about risk assessment. The issues related to biological research on microwaves and to risk assessment should be discussed openly.

ALLAN H. FREY

Randomline, Inc.
Huntingdon Valley, Pa.

To the Editors:

Foster and Guy ["The Microwave Problem," SCIENTIFIC AMERICAN, September] would have us believe that the only remaining questions about the safety of microwave radiation involve the irreducible uncertainties associated with proving a negative: "exposure to low levels of microwaves cannot be proved free of hazards." That is simply not true.

Research has barely scratched the surface of the issue. Of the 6,000 studies cited by Foster and Guy, only *one* has ever attempted a lifetime exposure to power levels incapable of producing a thermal response, and the results of that study are hardly reassuring. Nor has a single epidemiological study of a well-defined group exposed to nonionizing radiation ever been completed.

The one chronic study was carried out in Guy's laboratory with funding from the U.S. Air Force School of Aerospace Medicine (it served as one of the authors' examples). The study found a statistically significant increase in cancer among the rats exposed to microwaves.

Contrary to what Guy reports about his own results, the malignant tumors were heavily concentrated in the rats' endocrine systems: nine in the exposed animals compared with only two in the controls. (The total tumor numbers were 18 and five respectively.) In addition there were six pheochromocytomas—benign adrenal tumors—in the exposed rats but none in the controls.

That a principal investigator for a $4.5-million study should obscure his own positive results points to a basic feature of the "microwave problem": the domination of military funding for biomedical research on nonionizing radiation and the reliance on engineers rather than biologists to do the research. Indeed, Foster and Guy are both engineers.

The proposition that microwaves can have nonthermal effects is now widely accepted, even though it lacks a mechanistic explanation. Until we understand how microwave radiation interacts with living systems, the public will continue to be suspicious and the microwave problem will continue to be with us.

LOUIS SLESIN
Editor, *Microwave News*

New York, N.Y.

To the Editors:

Frey's ad hominem arguments are regrettable. If he disagrees with us on the scientific issues we raised (which included the nonreproducibility of some of the effects he has reported), he should prove us wrong.

One of us (Guy) has served as an expert witness. The role of an expert witness is to provide the best available scientific information in a lawsuit. In many of these cases Guy has consulted on behalf of plaintiffs suing industry, including companies that have sponsored his own research. Both of us have received research support from the military as well as from many other sources. Frey also has testified in such cases, and for many years his studies were supported by the military. Neither factor is pertinent to the scientific issues we raised.

The ANSI microwave standards do not reflect limited points of view. The committee that developed the 1982 standard had 18 members from the public sector and universities, 10 from industry, 12 from laboratories of Government agencies that are not major users of microwaves, such as the Environmental Protection Agency, and 11 from Government agencies that are major users, such as the Department of Defense. For several years drafts of the standard were circulated widely for comment. A different committee adopted the standard, and its members represented organizations as diverse as the Department of Defense and the Consumers Union of the United States. Other expert panels have independently arrived at conclusions similar to those of the ANSI committee. All these judgments reflect a consensus of informed opinion about a scientific literature that is in many places highly inconsistent and speculative. If some important hazard has been overlooked (and Frey does not indicate what it might be), the error is unlikely to be due to simple institutional bias.

The significance of the finding of increased cancer in the irradiated versus control rats is less clear to us than it apparently is to Slesin. The study involved 155 comparisons between exposed and unexposed rats. With so many comparisons, seemingly striking differences are expected by chance alone. It is misleading in such cases to ignore many negative findings and focus attention on one striking result. Slot machines encourage this error: no bell rings when the player loses money. Reports of "clusters" are similarly difficult to interpret. Interested readers might read an article by one of us (Foster) in *American Scientist,* March/April, 1986.

Certainly the probability was small (.005) that the greater number of malignant tumors in the exposed rats arose by chance alone. In a single comparison scientists usually consider an observed difference to be "statistically significant" if the probability of its having arisen by chance alone is less than .05. But given 155 comparisons, it is expected that one difference will be significant at the .005 level even though the two groups being compared are drawn from the same population. Moreover, the rate of cancer in the exposed rats was increased only in comparison with that of the unexposed controls and was about equal to the rate that is typically reported in such animals.

Therefore the finding is ambiguous. To build a strong case that exposure to microwaves is associated with the development of cancer in rats, follow-up studies would be needed that were designed in advance to test the hypothesis. That disturbing questions can be much more easily raised (by screening experiments or exploratory research) than settled (by studies testing hypotheses) is one of the long-standing problems in the microwave debate, and no doubt in other environmental health issues as well.

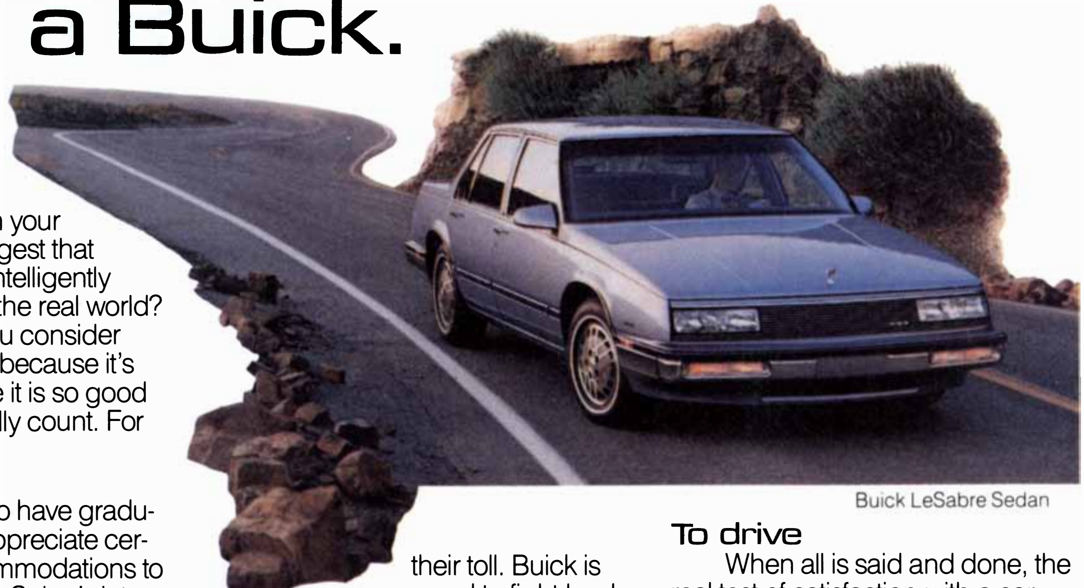
KENNETH R. FOSTER

University of Pennsylvania
Philadelphia

ARTHUR W. GUY

University of Washington
Seattle

If the world were a perfect place, you wouldn't need a Buick.

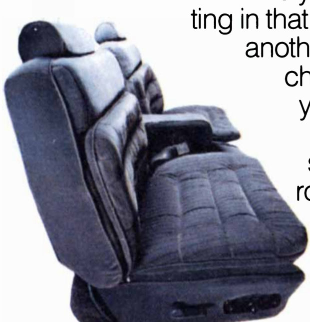


Buick LeSabre Sedan

The real world includes long stretches of traffic, terrible weather and serious demands on your car. So, may we suggest that you consider a car intelligently designed to handle the real world? We're suggesting you consider Buick LeSabre. Not because it's perfect, but because it is so good at the things that really count. For example...

For adults

Those of us who have graduated to adulthood appreciate certain significant accommodations to reality. A glance at LeSabre's interior tells you it's rich and handsome. Sit in it and you find something even better: true support combined with deep luxury. And the available 6-way power control seat puts a wide range of adjustments at fingertip control.



While you're sitting in that seat, do another reality check and you'll find another surprise: room. You have to stretch to reach the far side. Clearly, Buick is an automobile built for adult pleasure.

In the long run

The real world is not kind to automobiles. Rain, rocks, potholes, mud, salt, ice and humanity all take their toll. Buick is prepared to fight back.

Take something as fundamental as the paint. The LeSabre finish starts seven layers down. First, the body is immersed in a phosphate cleaning solution. Then it is electrically charged to provide a magnetic attraction for paint. Only then does the first primer coat go on. After the primer coat is wet-sanded by hand, a platoon of robots applies two separate layers of color enamel, dancing and whirling to cover every surface evenly and completely.

Finally, two coats of clear acrylic enamel are added to provide a deep, lustrous finish to keep your Buick handsome through years of life's abuses.

To drive

When all is said and done, the real test of satisfaction with a car happens behind the steering wheel. And Buick is just what the driver ordered.

LeSabre's big 3.8-litre V-6 with sequential-port fuel injection is there to do the job quietly, easily, reliably. Fully independent suspension not only smooths the ride but also grips the road in a highly sporting manner.

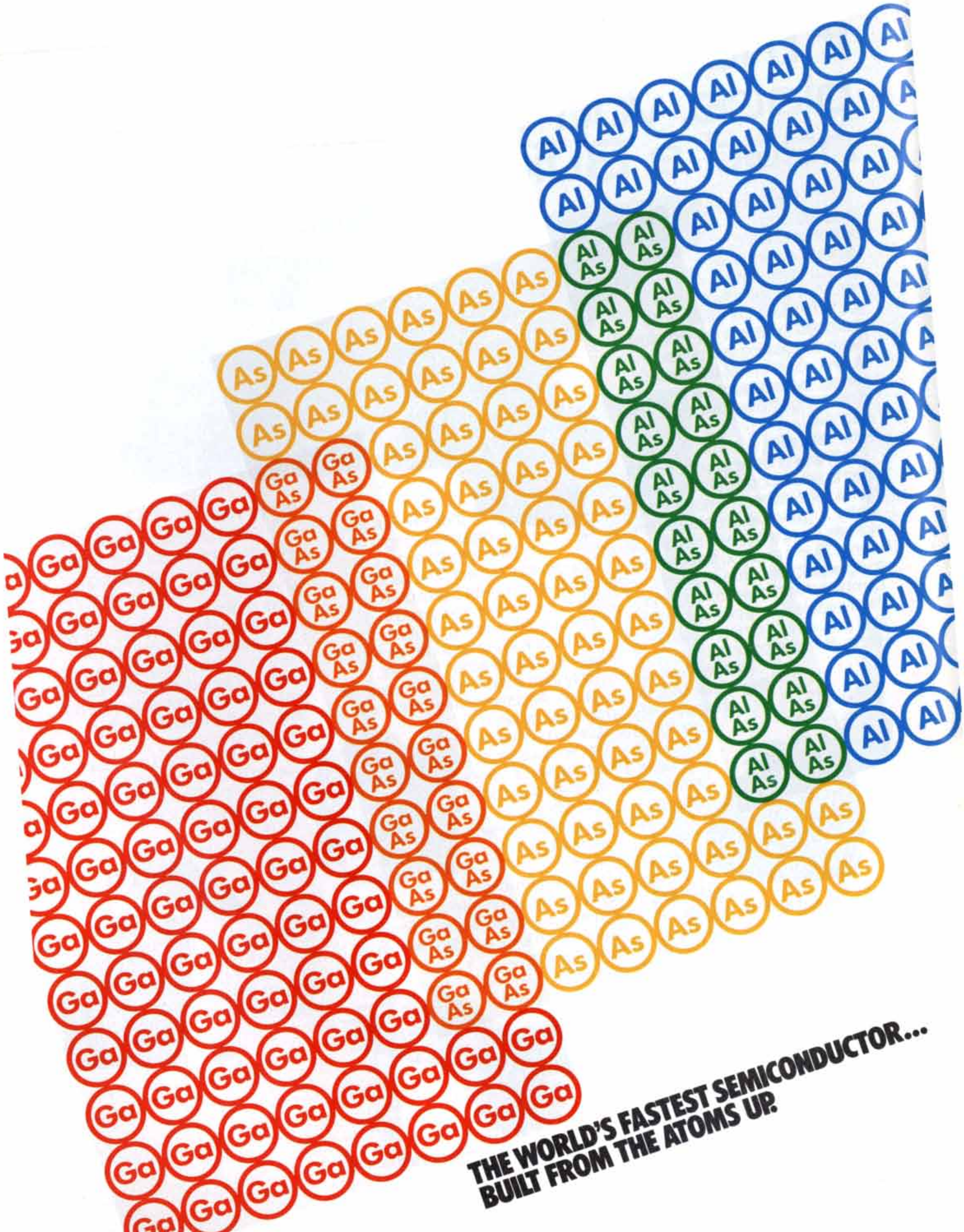
Until the world we live in perfects itself, may we suggest that you buckle up and see your Buick dealer for a test drive in an automobile that can make reality enjoyable. Really.

For more information, a brochure or a test drive, call **1-800-87-BUICK (1-800-872-8425).**

GM

Where better really matters.

B U I C K


THE WORLD'S FASTEST SEMICONDUCTOR...
BUILT FROM THE ATOMS UP.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

DECEMBER, 1936: "There is immense activity in preparation for the heavier-than-air conquest of the North Atlantic. There are, first of all, recurrent and apparently reliable reports that Pan American in conjunction with the British Imperial Airways is soon to put into operation transatlantic and New York–Bermuda airlines. Tests of the *Empire Flying Boat,* to be employed by Imperial Airways, have just been completed. A composite aircraft is now reaching completion. The upper component will be a two-float seaplane that is driven by four 350-horsepower engines. The lower component will be a large flying boat. Altogether there will be eight engines in use with a combined power of 5,000 horsepower. All the experts seem to agree that there will be no difficulty either in the launching of the composite aircraft or in the subsequent flying of the detached upper seaplane component."

"The production of about a dozen fruits from more than a hundred cross pollinations between different varieties of pumpkins and squash made at the State Experiment Station at Geneva, New York, has thrown considerable light on the botanical relationships of these groups and, incidentally, has given rise to several new forms that seem to be either immune or highly resistant to squash mosaic. Many unsuccessful attempts to hybridize these two vegetables have been made during the past century."

"A tiny instrument, weighing about 1¼ ounces and suspended beneath a small balloon, is science's latest device for finding the whys and wherefores of South Atlantic and Gulf Coast hurricanes. Signals for the release of the instrument-carrying balloons are given by the Weather Bureau. The three elements of the instrument record pressure, temperature and moisture. As soon as the balloon reaches its ceiling and bursts, a mechanism removes the penpoints from contact with the recording plate so that no records are made during the descent. A tag offers a reward for the return of the device to the Weather Bureau."

"At a cost of less than half a cent for each pint of water in which they are placed, cut flowers may now be kept fresher and more vigorous for longer periods of time before they wither and die. Experiments conducted at the scientific laboratories of the Hawaiian Sugar Planters Association in Honolulu indicate that adding two eyedropperfuls of sulfurous acid (not sulfuric acid) to each pint of water encourages buds to continue growing and leaves and stems to remain greener and permits the flower itself to retain its freshness, in some cases for days after it would normally cease to be attractive."

"Anyone who looks carefully at the Milky Way must be impressed by its irregular and patchy appearance. A generation ago it was supposed that the darker parts were gaps between the visible star-clouds through which the blackness of outer space was revealed. Then [Edward E.] Barnard, by a masterly analysis of his own superb photographs, convinced a skeptical world that the dark lanes and patches were caused by enormous opaque clouds—dark nebulae—lying between us and the bulk of the galactic stars, and blotting them out, or at least greatly diminishing their light."

SCIENTIFIC AMERICAN

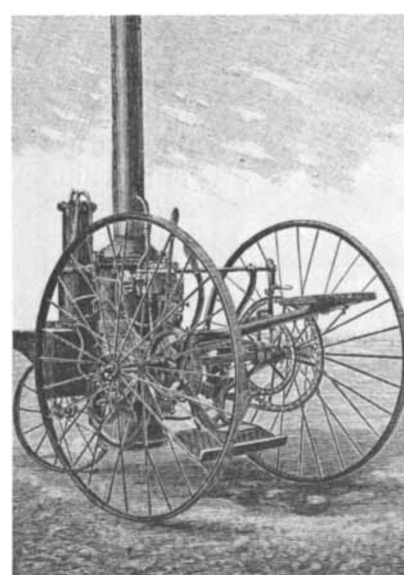
DECEMBER, 1886: "Diamonds have been objects of interest to all classes, but more especially to scientists and savants, to whom, even up to this present age, they are a mystery as to their origin or formation. Some attribute them to be of celestial origin, as aerolites, possessing electric light; others believe them of vegetable origin, since some are found with water cavities and also vegetable as well as animal matter embedded in them. Workers in them seem to have a more true and practical knowledge of them, and feel convinced that they are more of a volcanic origin."

"In Rochester, N.Y., on the 20th of November, 750 out of the 950 customers of the Bell telephone declined to use the instruments any longer, on account of the exorbitant charges. They are now casting about for instruments that can be supplied at cheaper rates."

"The *Manufacturers' Gazette,* we believe, speaks candidly when it says that the increasing use of opiates and other drugs intended to either allay or excite nervous activity is an evil in this country equal to if not worse than the excessive use of intoxicating liquors. Comparatively little is said of it in public journals, and there is no such crusade against it as there is against intemperance. The insidiousness of the drug habit makes it the more dangerous. The great majority of those who begin the use of opium, morphine and chloral do it under prescription of physicians, and often without being allowed to know what they are taking until the habit is thoroughly fastened upon them. Such trifling with life and health by physicians should be made a criminal offense, and its victim or his friends should prosecute for malpractice to the full extent of the law."

"Although the production of type writing machines, as an industry, at the present time has reached large proportions and obtained permanent footing among the great manufacturers, still it may be said only to be in its infancy. The utility of the type writer is so great, its success so marked, its applications so numerous, that no prophetic vision is required to perceive that, ere long, it will become spread throughout the civilized world, like the clock and the sewing machine."

"The question of terrestrial locomotion is one that interests a large number of readers. Mr. Louis Lallemand, a skillful mechanic in France, has just constructed a steam tricycle, to be heated by petroleum. Mr. Lallemand has already made four trials of this tricycle. On a level road a pressure of two atmospheres suffices to run the apparatus at a speed of from 3½ to 4½ miles per hour."



A petroleum-fueled steam tricycle


Significant Moments

THE AUTHORS

DONALD W. LIGHT ("Corporate Medicine for Profit") is professor of health care and psychiatry in the School of Osteopathic Medicine of the University of Medicine and Dentistry of New Jersey and professor of sociology at Rutgers University; he is also senior fellow at the Wharton School's Leonard Davis Institute of Health Economics. He holds a B.A. (1963) from Stanford University, an M.A. (1966) from the University of Chicago and a Ph.D. (1970) from Brandeis University. The research on which Light's article is based was done for the Twentieth Century Fund.

JAMES P. CRUTCHFIELD, J. DOYNE FARMER, NORMAN H. PACKARD and ROBERT S. SHAW ("Chaos") began their collaborative investigation of chaotic systems when they formed their own study group (called the Dynamical Systems Collective) as physics graduate students at the University of California at Santa Cruz. Crutchfield is a postdoctoral fellow at the University of California at Berkeley. All his degrees are from Santa Cruz. Farmer works at the Los Alamos National Laboratory, where he has been since 1982. He received his bachelor's degree in physics at Stanford University in 1973 and his doctorate from Santa Cruz in 1981. Packard is a member of the physics department as well as the computer-system research center at the University of Illinois at Urbana-Champaign. He got his Ph.D. from Santa Cruz in 1982, and he held positions at the Institut des Hautes Études Scientifiques in France and at the Institute for Advanced Study in Princeton before moving to Illinois. Shaw's current activities are not limited to the study of chaos: he is a pianist with a penchant for classical and improvisational music. He received his bachelor's degree at Harvard College in 1972 and his doctorate from Santa Cruz in 1980. While they were at Santa Cruz the authors applied their insights into dynamical-systems theory in attempts to beat the roulette tables at gambling casinos, but they learned firsthand how difficult it is to predict the states of random systems.

LAWRENCE M. KRAUSS ("Dark Matter in the Universe") holds a joint appointment as assistant professor in the departments of physics and astronomy at Yale University. He was an undergraduate student at Carleton University in Ottawa and did his graduate work at the Massachusetts Institute of Technology, where he obtained his Ph.D. in 1982. He was a junior fellow at Harvard University before taking his current position at Yale in 1985.

ROBERT C. GALLO ("The First Human Retrovirus") is head of a team at the National Cancer Institute studying the retroviruses associated with certain forms of cancer and with AIDS. He got his B.A. at Providence College in 1959 and his M.D. at Jefferson Medical College in 1963. After his internship he joined the National Cancer Institute, becoming chief of the Laboratory of Tumor Cell Biology in 1972. Gallo's team has done much to elucidate the role of viruses in the development of human leukemias and was one of the groups that first identified the AIDS virus.

RICHARD H. MASLAND ("The Functional Architecture of the Retina") is associate professor of physiology at the Harvard Medical School. He received his bachelor's degree at Harvard College and his doctorate from McGill University. Between 1968 and 1971 he was a postdoctoral fellow in neurophysiology at the Stanford Medical School, and afterward he moved to Massachusetts General Hospital. Masland set up his own laboratory there in 1973, and in 1975 he became a member of the faculty at the Harvard Medical School.

MICHAEL L. EVANS, RANDY MOORE and KARL-HEINZ HASENSTEIN ("How Roots Respond to Gravity") have brought both anatomical and physiological investigations to bear on the gravitational-response mechanism in plants. Evans, who obtained a Ph.D. from the University of California at Santa Cruz, is professor of botany at Ohio State University, where he specializes in whole-plant physiology. Evans writes: "As important as the techniques of molecular biology are for the study of plant hormones, it is essential that we not lose sight of the integrated function of the plant as a whole." Moore is associate professor of biology at Baylor University. His studies of plant responses to gravity arose from his interest in the anatomy of root cells. Moore's undergraduate education was at Texas A&M University and he did graduate work at the University of Georgia and the University of California at Los Angeles, where he got his Ph.D. in 1980. Hasenstein is a research associate in Evans' laboratory. He received two degrees from the University of Saarland in West Germany: a master's degree (1977) and a doctorate (1982). He spent two years as a postdoctoral fellow and adjunct professor of botany at San Diego State University before going to Ohio State.

ALEC N. BROOKS, ALLAN V. ABBOTT and DAVID GORDON WILSON ("Human-powered Watercraft") share an avocational interest in the application of human energy to various modes of transportation. Brooks works for AeroVironment, Inc., where his specialty in computational fluid dynamics has enabled him to develop computer simulations of bird flight. He studied at the University of California at Berkeley and at the California Institute of Technology, where he earned a doctorate in civil engineering. Abbott, who received his M.D. at the Indiana University School of Medicine, is assistant professor of clinical family medicine at the University of Southern California School of Medicine. He is currently studying the effect of exercise on people susceptible to coronary problems. Abbott has set speed records for human-powered land vehicles, some of which were achieved on streamlined bicycles of his own design and construction. Wilson is professor of mechanical engineering at the Massachusetts Institute of Technology. Born in England, he was educated at the University of Birmingham and the University of Nottingham, where he got his Ph.D. in 1953. He worked as a gas-turbine engineer in private industry and held various teaching and research positions before joining the faculty at M.I.T. in 1966. Wilson was made full professor there in 1971.

DONALD J. WATTS and CAROL MARTIN WATTS ("A Roman Apartment Complex") are respectively associate professor and temporary assistant professor in the departments of environmental design and architecture at Kansas State University. Donald Watts has a B.Arch. from the University of Nebraska and an M.Arch. from the University of California at Berkeley. Carol Watts is currently completing her dissertation in art history at the University of Texas at Austin. She received her bachelor's degree from Mount Holyoke College in 1971 and her architect's degree from the University of Washington in 1975. The husband-and-wife collaboration that resulted in the article began with Donald Watts's investigation of Roman cities in the Middle East and continued with Carol Watts's studies of Roman houses in Italy.

COMPUTER RECREATIONS

Of fractal mountains, graftal plants and other computer graphics at Pixar

by A. K. Dewdney

I can readily imagine the first full-fledged, feature-length motion picture generated by computer. The year is 1991. I stumble down the aisle while carrying an oversize bucket of synthetic popcorn and a soft drink containing a few additives that make all the usual ingredients unnecessary. The house lights dim, the curtains part and the silver screen comes alive with an adaptation of J. R. R. Tolkien's *The Lord of the Rings* trilogy. Frodo the Hobbitt strolls through an open glen. In the distance jagged, snow-capped mountain peaks thrust into the sky. In the foreground exotic trees and plants of unknown species shimmer in the sunlight. The scene changes to a wizard gazing into a crystal ball. In the center of the sphere a fortress appears, flames leaping from its battlements.

Although it is hard to say just how convincingly Frodo will walk and talk in such a film, I am convinced that the mountains, the plants, the crystal ball and the flames will all come off magnificently. The success will be due largely to the pioneering software and hardware of a company called Pixar, formerly the Lucasfilm Computer Graphics Laboratory. After visiting this fascinating hub of computer graphics in San Rafael, Calif., I am ready to share with readers the innermost secrets of mountains and trees. Anyone with a home computer can now generate images that closely resemble such objects. The space limitations of this department preclude an extensive treatment of crystal balls and fire, but I shall lay bare the basic principles of generating them.

In the hypothetical film described the camera might zoom in on the snowy peaks behind Frodo. Never would a more forbidding landmass be seen. Each large peak consists of smaller peaks, which are composed of still smaller ones, and so on: an infinite regress of peaklets. Even an Orc—that gargoylelike beast with leathery feet—would be uncomfortable standing on the rough slopes.

It is easy in principle to generate such a mountain range. For simplicity I assume that the terrain covers a triangular area. One then subdivides the triangle into four smaller triangles by finding the midpoint of each side and joining the new points by three line segments. Each triangle is then subdivided in the same way. The process is continued until the limits of resolution or of computation time are reached. The result—a rather boring lattice of triangles—can be enlivened by mixing in some vertical action: each time a new midpoint is added to the scene one displaces it upward or downward by a random amount. The random displacements, which must in general be decreased as the triangles become smaller, transform the triangles into crumpled peaks and folds [*see top illustration on page 16*].

Why should such a technique produce natural-looking mountains? The answer may lie in part with the fact that the process yields a fractal: a type of object that reveals more detail as it is increasingly magnified. Fractals are seemingly seen throughout nature. Benoit B. Mandelbrot, the indefatigable fractalist at IBM's Thomas J. Watson Research Center in Yorktown Heights, N.Y., uses coastlines to illustrate the basic idea. Imagine being asked to measure the French coastline with a measuring stick one kilometer long. One swings the stick end over end in a ponderous march along the shore and counts the number of kilometers. Many small bays and points are missed, however, so that the final length as measured in this way is not strictly accurate. Repeating the exercise with a meter stick produces a more accurate, longer measurement. But even in this case a large number of miniature embayments and spits are missed. No doubt a centimeter stick would be more accurate.

As a general rule, the measured length of the coastline increases as the measuring stick becomes smaller.. The relation between the measured length and the size of the stick is a special number. It is called a fractal dimension. A fractal dimension differs from an ordinary dimension in that it is usually expressed as a fraction, not as a whole number. The coastline in question might, for instance, have a fractal dimension of 3/2. Such a shape can be thought of as intermediate between a shape that has one dimension (a straight line) and a shape that has two dimensions (a plane). If a coastline were relatively straight, its fractal dimension would be close to 1; if a coastline were very rough, its fractal dimension would approach 2 as it attempted to fill a two-dimensional plane.

The fractal model of nature implies an infinite regress of detail. From the perspective of computer graphics the question of infinite regress is a red herring; it is enough if the landscape appears to have detail at all levels of magnification. Up to the limits of screen resolution the mountains to be generated have features as fine as the final triangles used in the subdivision described above. Although the complete mountain-drawing algorithm is too long and involved to describe here at a useful level of detail, there is a simple program called MOUNTAIN that draws Mount Mandelbrot in cross section. MOUNTAIN illustrates the essential idea of randomly displacing subdivision points along a vertical axis. The fractal artist begins with a single, horizontal line segment. The midpoint is determined and deflected up or down by a random amount. Each of the two resulting line segments is then subdivided and perturbed. The process can be continued in a manner analogous to the triangle-subdivision technique.

MOUNTAIN maintains two arrays called *points* and *lines* to keep track of the mountainous profile. Each array has two columns and enough rows (say 2,0 48) to accommodate handily one's screen resolution. The two columns of *points* contain coordinates and the two columns of *lines* contain indexes; each line is specified as the pair of positions in the array *points* that designate the coordinates of the line's endpoints. Because it is interesting to watch the successive subdivisions form a mountainous outline from an unpromising polygon, MOUNTAIN puts each generation under the user's control. At the end of a single, main loop the program asks the user if another iteration is wanted. If the answer is yes, execution branches back again to the head of the program.

The main loop converts the current

sets of points and lines into new sets that are twice as large. To do this it scans *lines* one row at a time, looks up the indexes of the corresponding points and retrieves their coordinates from *points*. Armed with the coordinates of a given line's endpoints, the program computes the coordinates of the line's midpoint, altering the y coordinate randomly in the process. The following algorithmic listing provides an adequate basis for a program. In it the variables j and k point to the rows of *points* and *lines* that are currently being filled with the latest results of subdivision. The variables *pts* and *lns* record the numbers of points and lines making up the mountain before the main loop is entered. Initially j is equal to *pts* and k is equal to *lns*. The index i runs from 1 to *lns*.

$$
\begin{aligned}
&j \leftarrow j + 1\\
&k \leftarrow k + 1\\
&a \leftarrow lines(i,1)\\
&b \leftarrow lines(i,2)\\
&x1 \leftarrow points(a,1)\\
&y1 \leftarrow points(a,2)\\
&x2 \leftarrow points(b,1)\\
&y2 \leftarrow points(b,2)\\
&points(j,1) \leftarrow (x1 + x2)/2\\
&points(j,2) \leftarrow (y1 + y2)/2 + random(range)\\
&lines(i,2) \leftarrow j\\
&lines(k,1) \leftarrow j\\
&lines(k,2) \leftarrow b
\end{aligned}
$$

This part of MOUNTAIN is largely self-explanatory. When the coordinates of the jth point have been computed, the index j is stored as the second point of the ith line and the first point of the kth line. The first point of the ith line is the same as before and the second point of the kth line is identical with the original second point of the ith line, namely the one with index b. When the loop is finally computed, *pts* and *lns* must be reset to the latest values of j and k respectively. The variable *range* is initially set by the user as the maximum amount of vertical randomness that can be given to the subdivision point. Each time the loop is completed this variable must be divided by 2 in order to keep the random fluctuations in scale with the size of the features being varied. The function *random*(*range*) is intended to express the selection of a random number between 0 and the current value of *range*.

If the mountains behind Frodo seem impressive, the trees and plants that surround him are no less so. They are both realistic and eerie. They seem real because they have branching patterns similar to those of actual plants. They seem eerie because they are not familiar species; the graphic designer has so many parameters available that he or she cannot resist the temptation to create something new.

The new "species" are called graftal plants, because they are based on graphs and have an implicit fractal nature. By implicit fractal nature I mean that the rules for generating the plant's basic topology could be (but are not) applied to the limit of resolution of the screen. In short, a twig does not regress indefinitely into twiglets. Once the graph underlying a plant has been developed, it can be converted into a myriad of convincing species by interpreting the graph in terms of size, color, thickness, texture and so on.

The graphs that underlie a given plant are produced by L systems, a class of grammars introduced by the Danish biologist and mathematician Aristid Lindenmeyer in 1968. An L system is essentially a set of rules for deriving new strings of symbols from old ones. The rules involve substituting sequences of symbols for single symbols. For example, using the numbers 0 and 1 and the symbols [and] one can generate a wide range of complex botanical forms with the following rules:

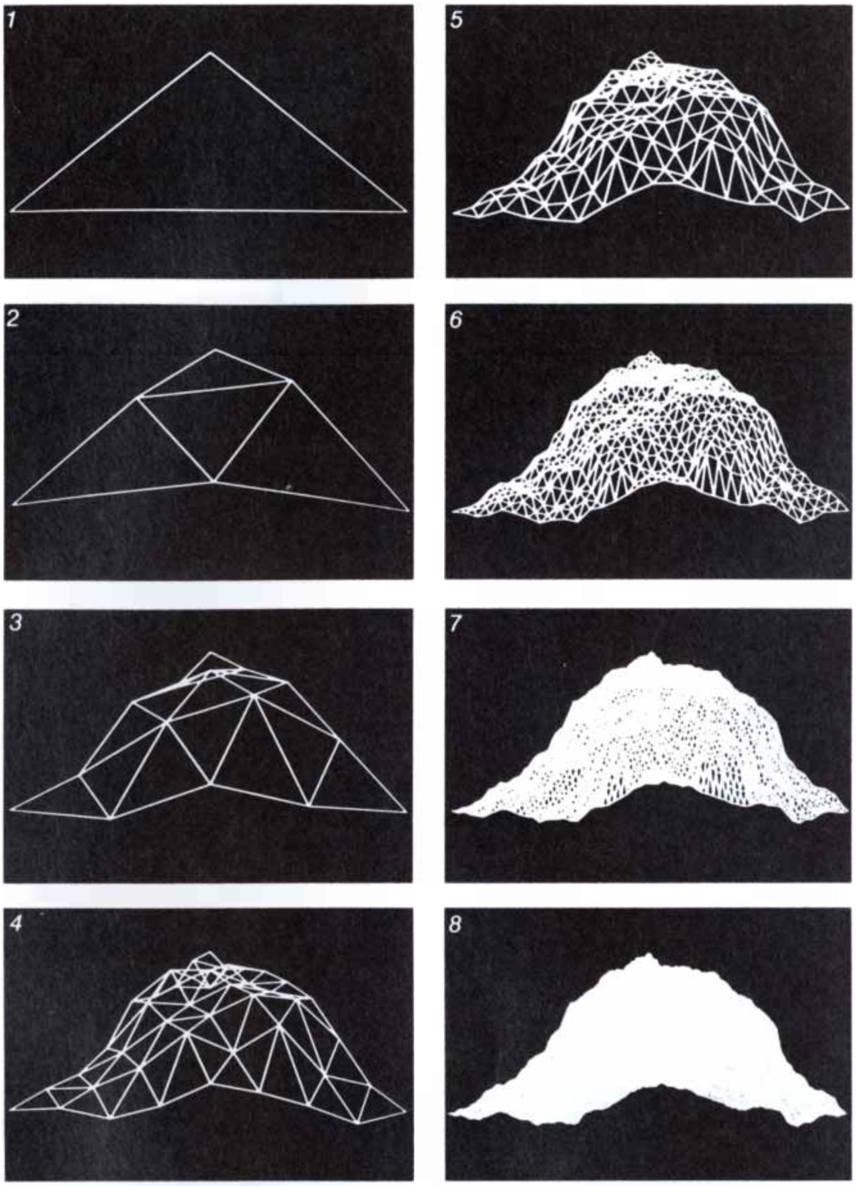
0 → 1[0]1[0]0
1 → 11
[→ [
] →]

To see how the rules work, suppose one starts with the string consisting of a single 0. For each left-hand symbol from the list one substitutes its corresponding right-hand symbol to obtain the following strings in succession:

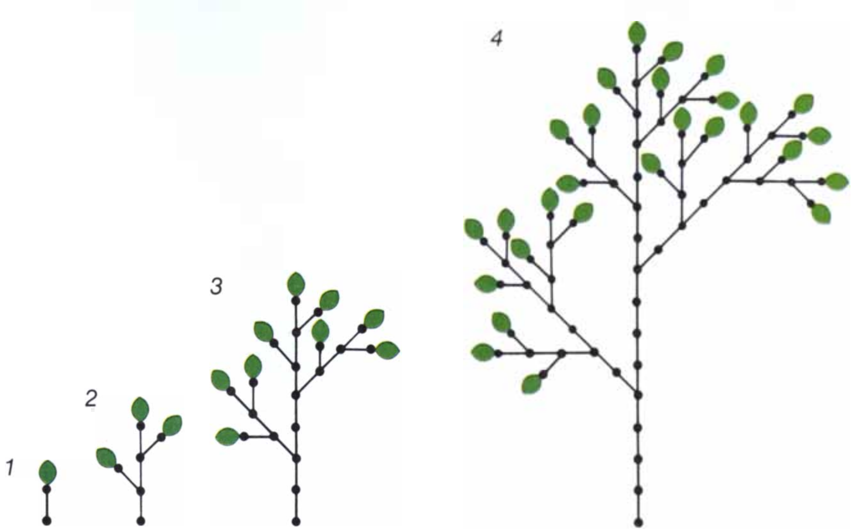
0
1[0]1[0]0
11[1[0]1[0]0]11[1[0]1[0]0]1[0]1[0]0

Such strings can be converted into treelike graphs by treating each num-

Computer-generated fractal mountains imitate nature



Subdividing triangles makes a mountain



Four generations of a graftal plant

ber (0 or 1) as a line segment and each bracket as a branch point. Both 0 and 1 segments are equal in length; they are typically distinguished from each other by making all 1 segments bare and placing a leaf at the outer end of each 0 segment.

The stem (or trunk) of the 1[0]1[0]0 string, for instance, consists of the three symbols not in brackets; a 1 segment is surmounted by a second 1 segment and topped off with a 0 segment. Two branches, each consisting of a single 0 segment, sprout from this formula. The first branch is attached above the first segment and the second is attached above the second segment. Before studying the bottom illustration at the left, readers might enjoy drawing the first few generations of the structure. For the sake of realism, additional interpretive features can be added to the model. One might specify that for any given stem (irrespective of whether it is the main one or not) the branches should shoot off alternately to the left and right. Not wanting to impose anything more arduous on those readers who would like to program graftals, I am happy to suggest mere stick figures for plants. The professionals at Pixar convert the grammar just described into beautiful plants such as the ones in the top illustration on page 18.

A two-part program called PLANT generates the *n*th string in the sequence above and then renders it as a line drawing. In its first phase PLANT maintains the strings it generates in two symbol arrays known as *stringA* and *stringB*. Each generation of plants occupies one of the two arrays in an alternate manner: the generation in one array is derived from the previous generation in the other. It is not strictly necessary to store symbols in these arrays; the numbers 0, 1, 2 and 3 will do nicely as long as the program substitutes correctly.

The *L*-system rules are embodied in conditional statements. For example, the following bit of algorithmic code can be adapted to convert a 0 in the *i*th position of *stringA* into nine new symbols in *stringB:*

if $stringA(i) = 0$, then
  $stringB(j) \leftarrow 1$
  $stringB(j+1) \leftarrow 2$
  $stringB(j+2) \leftarrow 0$
  $stringB(j+3) \leftarrow 3$
  $stringB(j+4) \leftarrow 1$
  $stringB(j+5) \leftarrow 2$
  $stringB(j+6) \leftarrow 0$
  $stringB(j+7) \leftarrow 3$
  $stringB(j+8) \leftarrow 0$
  $j \leftarrow j+9$

Here 0 and 1 stand for themselves

PATENT #4,584,240 IS THE ONE FISH STORY YOU HAVEN'T HEARD BEFORE.

On August 1, 1985, DuPont came up with a fish story that's so unbelievable we patented it. Patent #4,584,240 is Prime,® the first and only cofilament fishing line. The inner polyester filament is for low stretch. The outer nylon filament is for flexibility.

The result? A level of sensitivity, hooksetting speed and power that no other line can match.

Although DuPont has shown the way in nylon fishing line from the beginning, it isn't our only sports interest. DuPont products have brought new standards of performance to tennis, skiing, boating, hunting, swimming, golfing and bowling.

You see, we've earned more than 80,000 patents in the last 50 years. Some save lives, travel into space or help crops grow. Others, like Patent #4,584,240, have a different role. They just make your life a little more enjoyable.

BETTER THINGS FOR BETTER LIVING.

DU PONT

REG U.S. PAT & TM OFF

while 2 and 3 stand for [and] respectively. If the ith symbol of *stringA* is 0, then the program installs the sequence 1,2,0,3,1,2,0,3,0 in nine consecutive positions of the array *stringB* starting at index j (the first position of the latter array that has not yet been filled). A single loop in the first phase of PLANT contains four such conditional statements, one for each possible symbol encountered. The loop uses the index j as a reference to the symbol of the present generation currently being processed. The loop is executed for as many generations as the graftal artist wishes. At each stage PLANT may question the user whether or not another (longer) string of symbols is wanted.

The second, or graphic, phase of PLANT converts the string produced by the first phase into a drawing. It does this recursively. As long as no left bracket, or 2, is encountered, it draws a sequence of line segments in a given direction. When a left bracket of a given pair is scanned, the program draws the next line segments in a new direction, 45 degrees counterclockwise from the previous one. The end of the procedure is signaled by the appearance of the corresponding right bracket; here a leaf (whose shape and color I leave entirely to the reader's whim) can be drawn. The appearance of a second left bracket causes the process to repeat, except that now the new direction is 45 degrees clockwise. The rest is automatic.

PLANT uses a scaling factor that depends on the complexity of the plant being drawn. The nth generation, for example, is approximately 2^n segments high. If the screen is 200 pixels high, segments must be shorter than $200/2^n$. Ambitious readers will undoubtedly attempt variations in the generating grammar, branch angles and leaf shapes. If such variations are run on the same screen, landscapes of plants and trees (not very realistic, admittedly) will appear.

The crystal ball in the hypothetical Tolkien film would be made by means of a technique called ray tracing; the flaming battlements would be simulated by tracking the motion of a large system of particles.

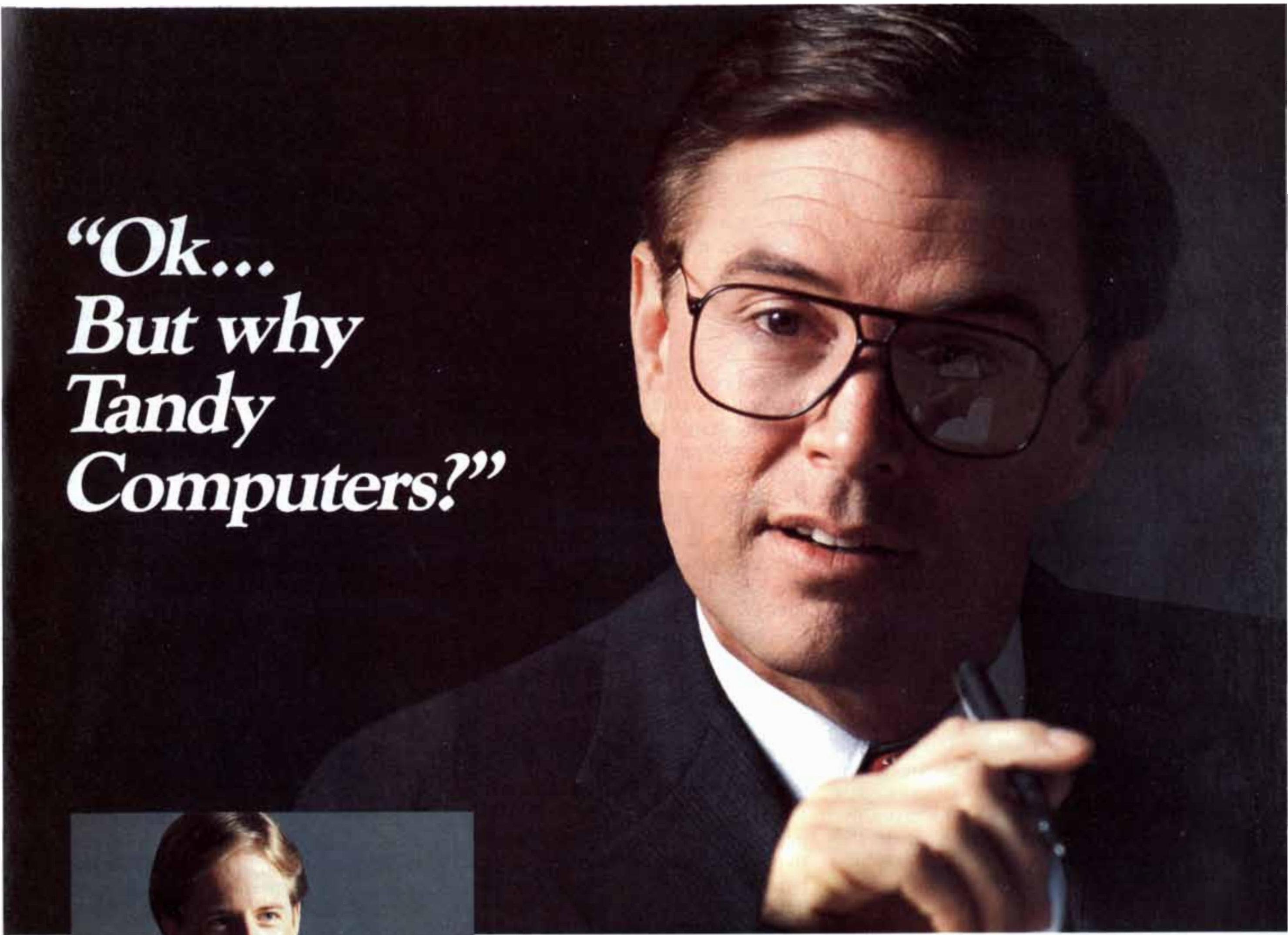
Ray tracing requires one to specify both the three-dimensional geometry of a scene and the position of a light source. When light leaves a source, it embarks on a complicated history of reflections and refractions. The eye of an observer standing in the scene will intercept some of the rays of light streaming from the source but will miss many—in fact most—of the others. To save computing time and power the ray-tracing technique works in the opposite direction. Imagine for a moment that light leaves the eye instead. A wide bundle of rays fans out into the scene. If a ray strikes a reflecting or refracting surface, it zooms off in a new direction determined by the laws of optics. Ultimately the ray will strike an absorbing surface, and it takes on whatever color is assigned there. That color is recorded at the pixel corresponding to the direction of the original ray.

An example of an image generated by the ray-tracing technique is shown at the left. A relatively simple history is experienced by the rays traced in the billiard-ball scene. On the surface of the balls we can nonetheless see reflections of the interior of a poolroom and a man holding a cue as he stands and watches the shot.

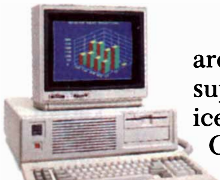
The large system of particles that might be used to generate the flaming battlements is a logical outgrowth of the little clouds of dots that symbolize miniature explosions in video and computer games. A particle system at Pixar is much more sophisticated than such chicanery, however. Within a certain region a host of particles live, move and have their being. Under computer control, each particle is a point that moves according to prescribed dynamics. Born at some time, it is allowed to move for a while, per-

Graftal plants from Pixar

A computer-generated image of billiard balls illustrates the ray-tracing technique



"In a word...quality."


Tandy 3000 HD

Tandy® computers are designed, built, supported and serviced by Radio Shack. Our total commitment assures a high degree of product quality and, just as important, customer satisfaction long after the sale.

Quality-built in the USA. Our MS-DOS® personal computer line is made in America in our own manufacturing plants. We control quality, from initial assembly to completed and tested product.

Nationwide Support. Over 1200 Radio Shack Computer Centers are ready to help you choose the latest in computer technology, backed by quality support.


Tandy 3000 HL

Training and specialized software instruction is available at your place of business or ours, in 60 major market areas.


Tandy 1000 SX

We've got what it takes. Like commercial leasing plans. And service performed by technicians who work for the same company that manufactured and sold you your computer.

You expect quality . . . and you get it from Radio Shack.

TANDY COMPUTERS:

In Business...for Business™

MS-DOS/Reg. TM Microsoft Corp.

Radio Shack®
COMPUTER CENTERS
A DIVISION OF TANDY CORPORATION

haps even giving birth to new particles. Then it may die.

Particle systems were spectacularly featured in a scene from the film *Star Trek II—The Wrath of Khan*. A Genesis bomb is dropped on a dead, cratered planet. The bomb creates a ring of strange, sparkling flames that eventually engulf the planet. When they finally die out, we see the surface transformed into a lush biosphere. The effect was produced by Pixar in its earlier incarnation as the Lucasfilm Computer Graphics Laboratory. The expanding ring of fire consisted of particle systems in which some particles gave birth to entire new systems. The new systems featured particles shooting upward from the planet's surface, changing color and even falling back again under the influence of gravity.

Alvy Ray Smith, who directs research and development at Pixar, took me on a tour of the company when I visited San Rafael. Smith is familiar to *Scientific American* readers for his work with cellular automata [see "Mathematical Games," February, 1971, and "Computer Recreations," August, 1985]. In addition to Smith, who pioneered the graftal approach to computer plant life, I met Loren Carpenter, who specializes in fractal mountains, Robert L. Cook, an expert in ray tracing, and William Reeves, the originator of particle systems. In the middle of a discussion about graphics software, Smith surprised me by saying that the company's main business is not so much the production of special effects for Hollywood as it is the manufacture of a special-purpose graphics computer called, naturally enough, the Pixar Image Computer.

At the heart of the Pixar Image Computer is a 24-megabyte, 2,000-by-2,000 pixel memory. That is more than enough resolution for most applications. Each pixel, moreover, is represented by 48 bits of memory, enough to store copious information about color and transparency. The Pixar's massive memory is managed by four parallel, high-speed processors that are fully programmable. They can execute approximately 40 million instructions per second, a speed several orders of magnitude greater than that of ordinary computers. A video-display unit communicates with memory at a speed of 480 million bytes per second.

The Pixars were first shipped this May. They are destined for careers in medical imaging, remote sensing, engineering design and animation. Perhaps they will even be used to generate my hypothetical film.

Wallpaper for the mind, the subject of this department in September, treated computer images that are almost but not quite repetitive: a kind of wallpaper not yet seen. In fact, however, patterns resembling these were known a generation ago. Michael Rossman, a writer and political commentator living in Berkeley, Calif., coined the phrase "wallpaper of the mind" in 1971. He was referring to the delicate patterns seen with the eyes closed after the ingestion of LSD. Rossman writes: "It 'looked' like wallpaper, never striped but of repetitive motif: orderly constellations of parrots, starfish, lightbulbs, snowflakes, unnamable Rorschachs, changing in profligate creativity....the images evolved one from another in chain, as if some infinite linear Escher print of metamorphoses had been animated in a movie."

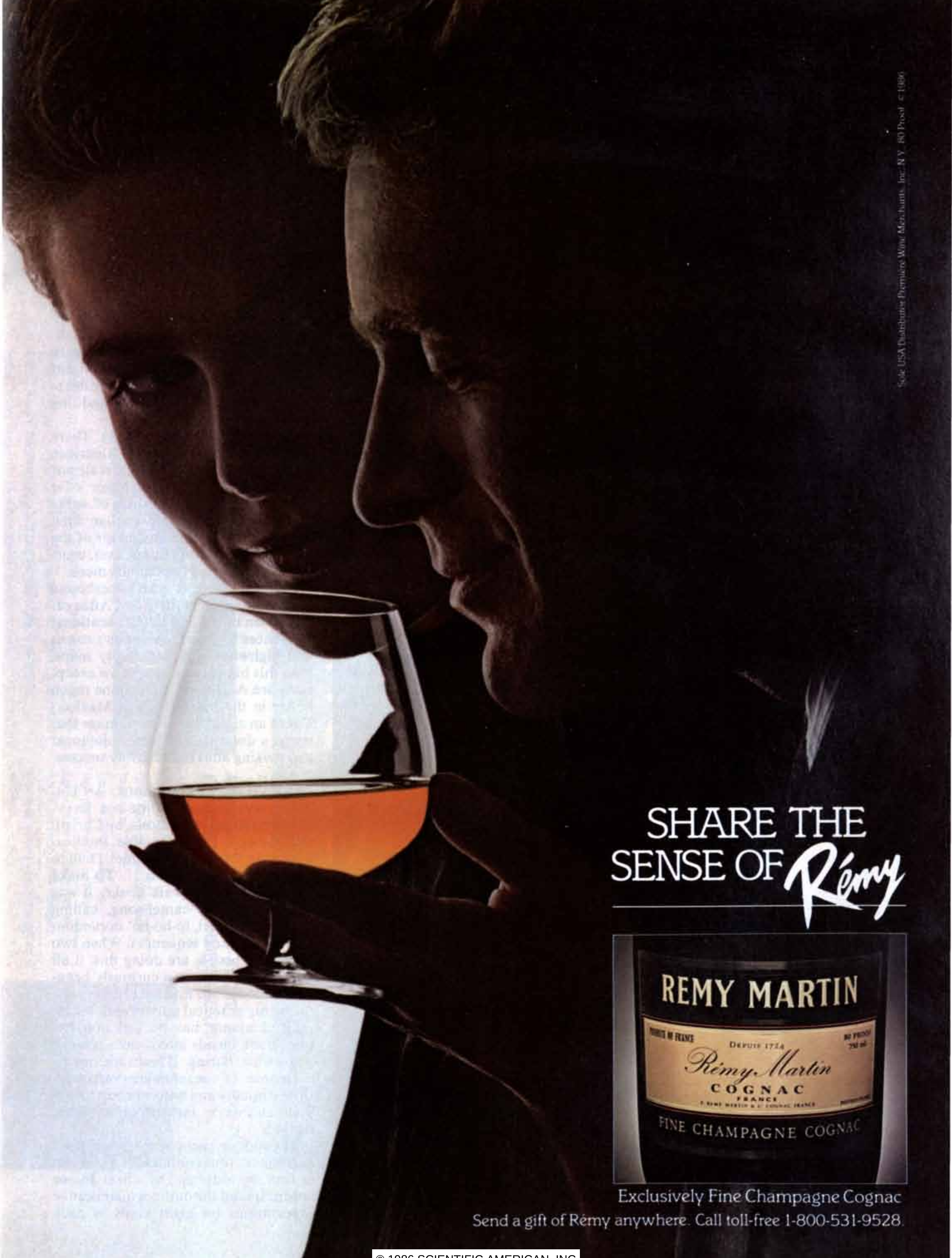
Readers will remember that the simplest wallpaper program described in the column belonged to John E. Connett of the University of Minnesota. According to his prescription, the coordinates of each pixel are squared and added together. The result is truncated to an integer. If the integer is even, the pixel is colored black; if it is odd, the pixel is colored white. The appearance of wallpaper—horizontal and vertical repetitions—may be explained in part as a moiré phenomenon: two patterns are implicit in the foregoing description and their superposition creates the effect. The first pattern is the rectangular grid of pixels that constitute one's display screen. The second pattern is a series of concentric rings that represent points in the plane for which Connett's procedure generates an odd number. The wallpaper arises as a result of repetitive interference patterns in the horizontal and vertical directions. A number of consecutive grid points may happen to fall within rings, the next bunch will then fall without, and so on. As the distance from the origin increases, the rings become progressively thinner, ensuring that hits and misses will occur along any line of grid points.

As if to reinforce these remarks, Paul Braun of Simi Valley, Calif., allowed impatience to get the better of him. It took so long to fill his computer screen that he decided to sample the pattern by displaying every ninth pixel in both the horizontal and the vertical direction. In compressed form the resulting image looked nothing like the original; Braun had in effect changed the grid size and thus altered the interference pattern.

Otto Smith of Port Townsend, Wash., found he could change the moiré pattern considerably by merely varying the color scheme. Like many readers, Smith selected formulas other than Connett's. Certain sums or products of simple trigonometric functions, for example, produce wild, swirly images reminiscent of the colored-ink patterns that used to adorn the inside covers of quality hardbound books. Smith also points out that moiré patterns and other interference effects crop up regularly in computer graphics. In such a context they are known as aliasing: the tendency of unwanted imagery artifacts to appear when a picture with much regular variation in it is digitized.

I was touched to receive a miniature hand-held film made by Douglas W. Raymond of Orinda, Calif. It consists of consecutive, minute prints of Connett's wallpaper in which the grid size is gradually increased. The prints are sewn tightly together and the production unfolds as one riffles the edge of the resulting booklet with the thumb. Raymond calls the film "Small Bang with Aliases."

A scene from the Genesis sequence in **Star Trek II**


SHARE THE
SENSE OF Rémy
REMY MARTIN
DEPUIS 1724
Rémy Martin
COGNAC
FRANCE
FINE CHAMPAGNE COGNAC
Exclusively Fine Champagne Cognac
Send a gift of Rémy anywhere. Call toll-free 1-800-531-9528.

BOOKS

Deck the halls with new books about science for young readers

by Philip Morrison

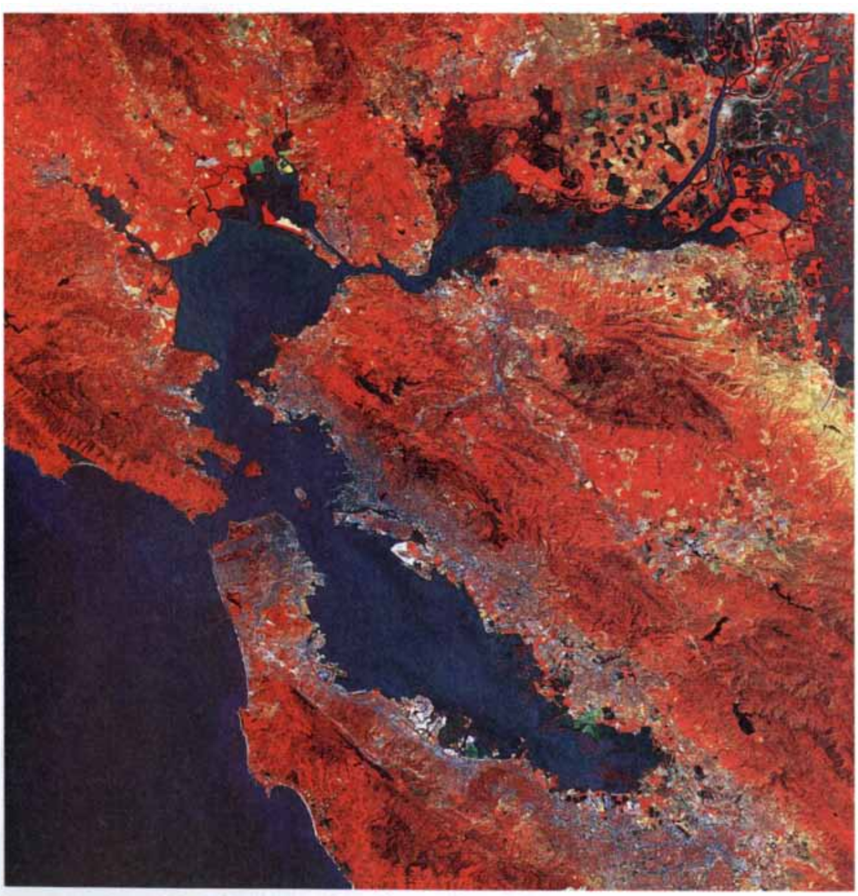
Maps and Field Guides

ATLAS OF NORTH AMERICA: SPACE AGE PORTRAIT OF A CONTINENT, Wilbur E. Garrett, editor. National Geographic Society ($39.95; flexicover, $29.95). About as many of the large pages of this new North American atlas are devoted to colorful video images from space as are given to the clearly printed political maps. The false colors of the video images—a few are even made to look rather natural—are the only barrier to viewing the picture as a photograph from above, an easy stance for a beginner to adopt. Yet the images are coded in color and displayed in a variety of scales. Each image was made at a particular time, perceived by a number of different sensors. Through such comparisons the familiar, more abstract drawn maps, necessarily crowded with text and symbols, acquire clear meaning. This is an inviting and instructive way for young readers to understand what a map does—and does not—represent.

The Southeastern states, for example, are conventionally mapped from Hatteras to Key West and Natchez, the towns and streams and roads clearly marked. The overleaf from that spread displays a nighttime radiometric satellite image, coded for temperature, which crimsons that same long coast with the Gulf Stream eddies of one November night. The map and radar image are followed by an Apollo close-up of the barrier islands just north of Hatteras, as well as Landsat photographs of the long, hazy ridges parallel to the Cumberland Mountains and of the Everglades and their mangrove fringe. Another page shows the Great Lakes in a big Nimbus image, deep blue within a green land; elsewhere the cities of New York and San Francisco appear from orbit. A Landsat photograph of the end of the Gulf of Honduras makes the nature of underdevelopment quite clear: the editors title the page "Coping in a hard land." A single road, its yellow dirt surface a long scratch across the deep-green forest, is the only overland link between two national capitals.

The touch is not always sure. There are many maps of North American cities, neither done to a single scale nor yet very detailed. Other pages offer state birds and flowers, lists of parks and public events, and similar data. More useful are thematic maps of the continent, indicating population, energy sources, geology and much more.

It is still true that a $6 paperbound annual Rand McNally Road Atlas offers more bread-and-butter locational references for North America's towns and highways, and better city maps, than this big volume can. (Two exceptions are Alaska, which is done much better in the big book, and Mexico.) Yet as an aid to the study of maps that opens a door wide to their enjoyment this striking atlas is a runaway success.



A Landsat image of the San Francisco Bay region

AN EXPLORER'S HANDBOOK: AN UNCONVENTIONAL GUIDE FOR TRAVELERS TO REMOTE REGIONS, by Christina Dodwell. Facts On File Publications ($14.95). "For my camel, I had to use Turkana commands.... To make him hurry up and walk briskly it was usual to sing a camel-song, calling 'Ha, ha, brrr; hei, hi-ho-ha' in random tones, notes and sequences. When two or several people are doing this, it all blends together into a curiously beautiful melody." In much the same way, the highly practical information the insightful author has packed into her brief book blends into a siren song of worldwide daring. There are strong harmonies of consummate craftsmanship, empathy and natural courtesy toward strangers, and there are happy endings.

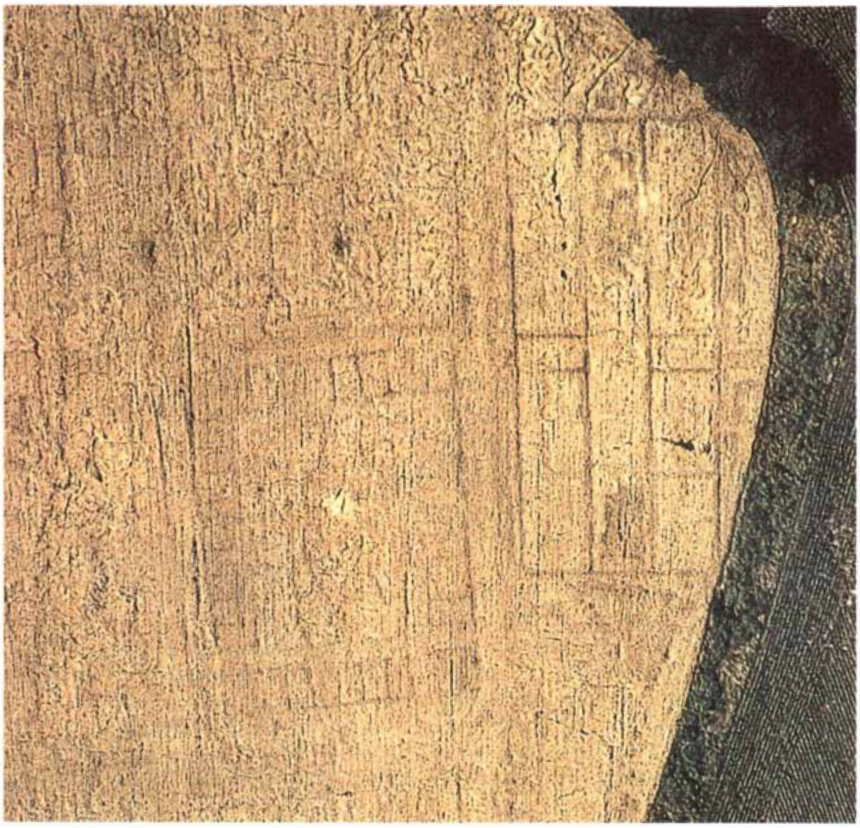
The author spent seven years wandering over four continents on two feet or four, by water and by wheel. In one cogent spread she outlines quantitative expectations for eight kinds of pack animals, from donkeys to camels. Elephants "are interminably slow walkers," and it may take months to earn your animal's trust. A long trip with horses can average 20 miles a day, at 200 pounds per horse. Camels carry three times that load a little more slowly. The author's tips include advice about getting out of tight corners with the human beings one encounters. Stay overnight in the village of people who seem hostile; they are honor-bound to protect guests, even if they might rob you in the bush. Country folk have the stricter codes of conduct; the remoter the region is, the safer—in general—is the traveler.

The second major topic is the choice of campsite, its fire, water and housekeeping. Fire making is treated with the care it deserves. It seems easy until you try; the craft lies in the arrangement of the fuel. Water is even more vital; in the desert it is training yourself not to drink casually that counts. "Drinkable water may be white, yellow, black, red, brown," for generally mud settles. Suspicions of hepatitis and bilharzia demand microbial vigilance by boiling or tablets.

Food receives close attention. There is an eloquent recollection of one desert gingerbread, baked of maize flour and camel's milk in embers glowing within a stone-lined hole, camel herdsmen nearby. "They sat down near my camp, but...they weren't watching me.... In a high nasal voice...he sang his strange falsetto song" to the horse. They had never seen a horse in this part of the world; the song was in reverence. The spicy loaf was shared. "Baking can be a wonderful thing."

No meat is wasted. Elephant trunk? Prepare it like beef tongue. Smaller heads are used in full, often baked whole in hot ash, after they are defurred and the mouth and ears are cleaned. Fruit bats are good to eat, along with field mice, all vegetable-eating rats of the bush, and many grubs and insects. Roast grasshoppers mix well with peanuts; curried locusts are good, although oily.

This book is a gleaming amalgam of adventure with hardheadedness, a find for good readers of any age who want to learn both to plan and to dream.



Traces of a Roman legion's camp, revealed in an aerial view of a wheat field near Vienna

THE FIELD GUIDE TO PREHISTORIC LIFE, by David Lambert and the Diagram Group. Facts On File Publications ($17.95; paperbound, $10.95). If it is not literally a field guide, this book, a bargain in paper, stakes a quite arguable claim. The experienced author and the London design studio have put together a first-rate beginning book in paleontology, one that marches presto across 120 spreads, each bearing a brief up-to-date text and many drawings.

A typical early spread shows the small silhouette of a man standing on the topmost of eight platforms, each broader than the one above it, each occupied by a number of forms representing all eight classifying categories from the superkingdom Eukaryota to the species *Homo sapiens*. A much later spread of placental pioneers shows small marginal sketches of platypus, wallaby and shrew; next to each animal are diagrammed the pelvic openings that distinguish the three mammalian groups. On the opposite page we see a simple reconstruction of one animal from each of five early orders of mammal. Other pages provide a chicken to give scale to huge flightless birds, and a light battle tank to help assess three plated dinosaurs.

The essential background is not slighted. Here are small maps showing the continental masses at various times, charts of the geologic periods (including the epochs of the Cenozoic), times of mass extinctions, the tree of five-kingdom life and simple accounts of the methods and goals of paleontologists. The center of the work consists of seven chapters that describe and display forms of the main groups over time. Plants and invertebrates get a chapter each, but the vertebrates take pride of place: chapters class by class, each with drawings and diagrams large and small, so that the look becomes that of a careful field guide. What a time it must have been in the late Miocene of North America, when graceful protodeers such as *Synthetoceras* chewed its coarse cud, and *Cranioceras* its better-processed one, both forms in full enjoyment of three horns each!

A few pages list and annotate the main museums in two dozen countries where interesting fossil collections are on display. (There are imposing Gobi dinosaurs to be seen in Ulan-Bator.)

THE PRACTICAL ARCHAEOLOGIST: HOW WE KNOW WHAT WE KNOW ABOUT THE PAST, by Jane McIntosh. Facts On File Publications ($18.95). Again a London production group has made a first-rate book, well conceived and executed, good for readers young and old. The text is the work of a Cambridge archaeologist, who has told an up-to-date and expert story of coming to know the past. She begins with a quick history of the science, describes the growth of methods for work in the field, particularly dating and mapping, continues with the essential work done after excavation has been completed, and closes with a survey of the theoretical issues of the day. A coda records with photographs the month-by-month progress of a dig in Billingsgate, City of London, in which John Schofield reports how they made their way patiently down through 10,000 layers from fish market to a Roman riverside quay.

It sounds like a fine textbook, and rather a personal one; on it has been placed a structure that makes it much more widely accessible, at the price of striking a certain staccato tone. The volume is organized within its large divisions topic by topic, one or two pages at a time, each page offering valuable pictures and diagrams, some in color. The topics include Egypt, Knossos, the grid technique of excavation and the extraordinary bulls'-head shrine at Çatal Hüyük. There are four fine pages on dating techniques and one on the origins of agriculture.

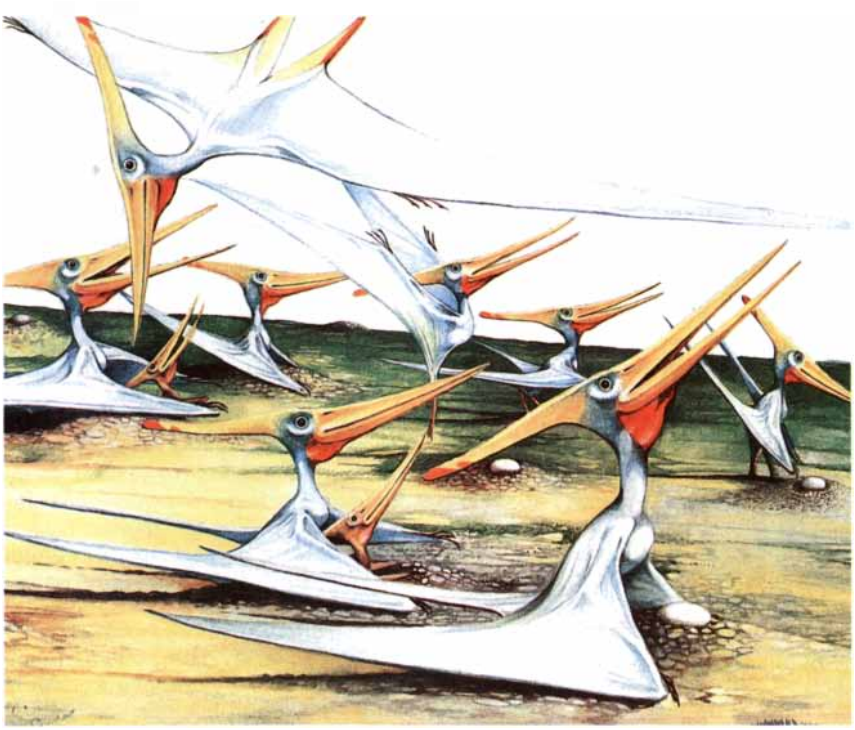
It is nonetheless not a ragbag; the intelligence and scholarship of the author keep the line of argument strong through the kaleidoscope of scenes. An encyclopedic quality is surely present, one that encourages more sustained attention. If a single spread on Mohenjo Daro and a Shang bronze vessel, both in color, can be taken for what it is and not for what it omits, the work is a success. It certainly brings value to any school library and to many a home, class or club.

ROADSIDE GEOLOGY OF NEW YORK, by Bradford B. Van Diver. Mountain Press Publishing Company, P.O. Box 2399, Missoula, Mont. 59806 (paperbound, $12.95, postage $1). The field geologist walks out the formations, hammer and sample bag in hand. The art is long. But a royal road to some grasp of the geologic picture exists for the purposeful motorist who views the same landscape less intimately, although at a comparable scale. What the touring family needs is the nontechnical guidance of a knowing geologist, with interpretation keyed carefully to the view from the highway, or sometimes from a vantage reached on foot.

This inexpensive guide to the exciting if complex geology of the Empire State was prepared by an expert geologist at the State University of New York at Potsdam. He and his wife have happily toured and logged almost 5,000 miles on the state's motor roads. Nearly 60 stretches of numbered highway are described here and anatomized in many photographs and suitable geologic maps; each page covers about 25 miles. The book draws heavily on plate tectonics, a narrative in which New York appears as a frontline state. A second theme is equally fundamental: the bones of the state, shaped in the deep past when the Atlantic was juvenile, are well clothed in most places by processes that ice much more recently set in motion.

The Niagara Frontier earns the first close account, geology caught in the act. A second exciting upstate tour circles and recircles the Finger Lakes, where the story of the glacial past is plain. Then there is the great dome of the Adirondacks, new mountain forms built out of ancient rocks, perhaps the work of a still rising plume. Downstate people have plenty to see as well, from Montauk to the Palisades. As one prize, consider the suspected meteor crater six miles across, to be viewed from state route 28 not far beyond Kingston!

This guide is one of an ongoing series. The publisher has managed so far to find enough expert and enthusiastic authors to treat half a dozen states and three mountainous national parks.



Nesting pteranodons, large late Cretaceous pterosaurs

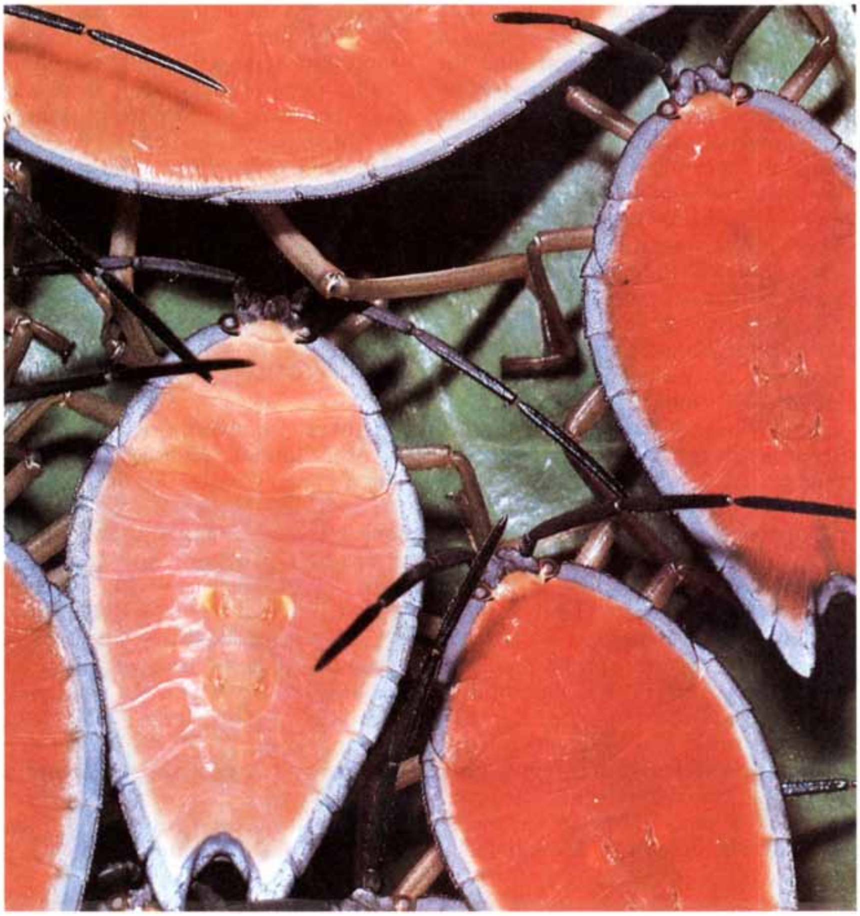
Biology

PTEROSAURS, THE FLYING REPTILES, by Helen Roney Sattler, illustrations by Christopher Santoro. Lothrop, Lee & Shepard Books ($13). WHAT THE ELEPHANT WAS: STRANGE PREHISTORIC ELEPHANTS, by Miriam Schlein. Illustrated with museum drawings, paintings and photographs. Atheneum ($15.95). THE DINOSAURS AND THE DARK STAR, by Robin Bates and Cheryl Simon. Illustrated by Jennifer Dewey. Macmillan Publishing Company ($11.95). These three excursions into paleontology differ in form and scope. They all present vivid reconstructions in word and picture that import into grade school science a delicious fear of monsters and dragons that cannot harm—certainly part of the timeless appeal of these creatures long gone. The books also do much more: all three offer something of the evidence and the reasoning that led to the imaginative results.

The freshest of the books offers a remarkable set of large paintings in plausible colors (what a lively reptilian chick looks out at you!); it treats the reptile group—not closely related to stegosaurus, brontosaurus and kindred species that ruled the skies during the Age of the Dinosaurs. A fine page shows the bones and imprint of the "shaggy devil," *Sordes pilosus,* exposed in 1970 on fine-grained golden stone by a Soviet researcher. Its entire body was visibly furry, and the warm-blooded nature of these small early fliers becomes plausible. When the artist gives the big *Pteranodon* white fur, we are told that the conjecture suits the seagoing 30-foot flier as camouflage against its surface enemies. Another toothy flier was surely a fisheater, because "fossil fish have been found in its rib cage." Toward the end of their time some reptilian fliers grew to remarkable size, as big as sailplanes. Here one ultralight form is reconstructed in the role of an enormous vulture, a couple of them scavenging the great carcass of a herbivorous dinosaur. They may have soared in thermals to seek food, and could probably lift off in a light breeze.

The second volume, illustrated with black-and-white versions of museum paintings of the old forms along with photographs and drawings (the jacket reproduces one famous mammoth painting in color), tells a similar tale about elephants. Once again the evidence spins a tale of wonder, although the morphological story is subtler; we cannot quite grasp here how the trunk is inferred from the nasal bones of the fossil skulls. The sequence, though, is clear and evocative; the four-tusked early elephants and a later form that held its trunk off-center from its closely spaced tusks are striking finds. A 16-foot tusk recurving toward the body of a huge woolly mammoth is a sight to behold.

The third volume tells clearly and concisely the celebrated tale of how the dinosaurs were done in (as diverse investigators maintain) by the consequences of periodic cometary collisions. Enough context is offered to make the story clear to grade school readers, and the grave doubts that surround the Nemesis hypothesis are not neglected. The drawings that decorate most page bottoms and margins are helpful, although the scale of one that shows the famous layer of clay at the end of the foraminifer-rich limestone in Gubbio is obscure. Even if the dark star is not there, its story is well told.



Clustered nymphs of the Australian shield bug

THE ENCYCLOPEDIA OF INSECTS, edited by Christopher O'Toole. Facts On File Publications ($24.95). The work originates in Oxford; the editor is an entomologist there and nearly all his 20 expert authors are from the two gray-stone universities or from the Natural History Museum in London. The uncommon good sense of the collaborators has allowed this slender reference volume to escape pedantry. In addition to insects the work sensibly discusses all the related arthropods that dwell on land, such as millipedes, spiders and scorpions. (Marine creatures of all leg counts are excluded.)

The plan is familiar: there are two dozen chapters, each with a brief, up-to-date text and a carefully chosen set of illustrations. A mixture of showy photographs and paintings in color evoke some authentic wonders; careful diagrams and close-set boxed text offer detail. The articles are concise and introductory in approach and serious in content. They neither avoid technical words and matter nor neglect important arguments.

The first few entries are incisive discussions of general importance. They touch on evolutionary history, chitin and exoskeletons, development, flight, pheromones and the like. The rest are accounts of groups of related orders, from bristletails to spiders.

Some samples picked out on a stroll through the rich pages will help to support this appraisal. Text and a picture of a generalized body plan for an arthropod celebrate the common wrapping of all these animals, the cuticle. A polysaccharide-and-protein composite, it is sometimes hardened by cross-linking (tanning), sometimes by mineralization. The termites are thoroughly described. They harbor pools of organized microorganisms, not only the famous gut protozoa but also many other bacterial and fungal forms with powerful biochemical abilities.

One photograph, all greens, light and dark, shows—if one looks hard—a stick insect from Venezuela that matches its aging leaf background by means of a set of lichen-mocking fanlike outgrowths. Such parasites as lice and fleas are well treated, and the 90,000 species of true flies are discussed with some affection, all the more unexpected since an accompanying section makes it clear that the two-winged flies are "by far the worst insect scourge of mankind." Some are flying syringes, some are flying contaminant smears, and some (the botflylike forms) deposit eggs which turn into larvae that feed within the open wounds they make in the living host.

This is a fresh adult reference in adult language, made unusually accessible to interested young readers by its brevity, visual appeal and topical organization into small parts. There are several companion volumes also.

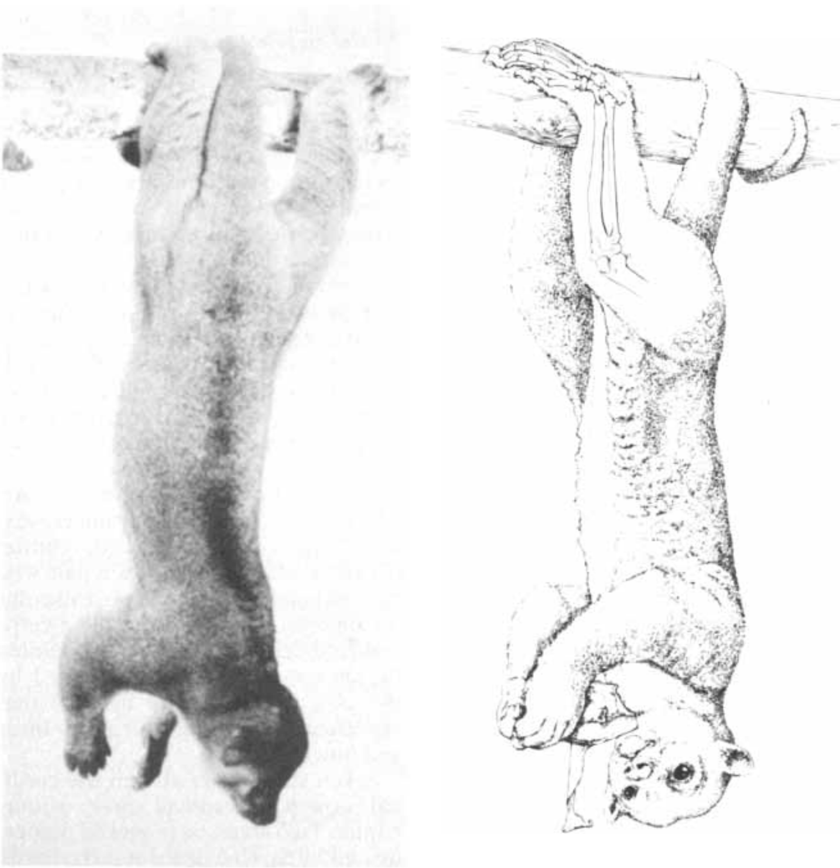
STATE BIRDS, illustrations by Arthur Singer and Alan Singer, text by Virginia Buckley. E. P. Dutton ($14.95). The father-and-son pair who here so handsomely depict in full color the birds in their habitats hold a certified authority: it is they who painted the designs for the stamps issued by the U.S. Postal Service in 1982 that represented the 50 emblematic birds and flowers.

Seven states have chosen the cardinal, now a widespread showy suburbanite. Two states have picked domesticated birds. One of them is the hardy Rhode Island Red that Isaac Wilbur bred and made famous at the turn of the century; few New England kitchens will as yet accept the white eggs of other stocks when they can find the decorous brown eggs laid by Red hens. The second is the Blue Hen's Chicken, the gamecock taken to camp by a company of sporting Delaware soldiers during the War of Independence, now more a tradition than a living bird. There is discriminating taste in Louisiana's choice of the brown pelican, ungainly in form but superbly graceful in diving flight, and in the selection by lake-dotted Minnesota of the common loon with its compelling call and wild laughter. Vermont named the hermit thrush, whose fluid song and literary prominence end this muster with distinction, 28 forms for 50 states. Birds have no need to fit state boundaries one by one.

IN THE POND, by Ermanno Cristini and Luigi Puricelli. Picture Book Studio USA, distributed by Alphabet Press, 60 North Main Street, Natick, Mass. 01760 ($9.95; paperbound, $4.50). Some dozen painted spreads, wordless, open, cool and colorful, display the life of a pond in summer, both above and below the surface: the probing beak of the stilt, the trowel of the mallard, the tree frog leaping for a dragon fly, the cattail, the inchworm, the kingfisher.... The pages form a panorama arranged so that the young reader—who needs no skill at all with text—can come to grasp the geography of this enclosed world. Two final black-and-white pages offer a key to the arrangement and to the 30-odd forms painted, and there is an opening half page of text. The delightful little book, of European origin, has two close kin made by the same artists, one of the woods and one of a garden.

Single Species

PHOEBE THE KINKAJOU, by Dorcas MacClintock, photographs by Ellan Young. Charles Scribner's Sons ($12.95). Nine bronzed furry pounds, long graceful tail and big eyes, Phoebe is a raccoonlike omnivore who moved with liquid grace through the high leafy canopy of the Panama rain forest. She was brought to Harvard by Deedra McClearn for a study of arboreal locomotion; kinkajous have swivel ankles, go down tree trunks headfirst and hang easily by hind feet they can reverse 180 degrees. Once her three years of clambering work was finished, the expert author-naturalist adopted Phoebe as a live-in guest.



Phoebe the kinkajou, hanging from a tree limb by swiveling her hind feet **(right)**

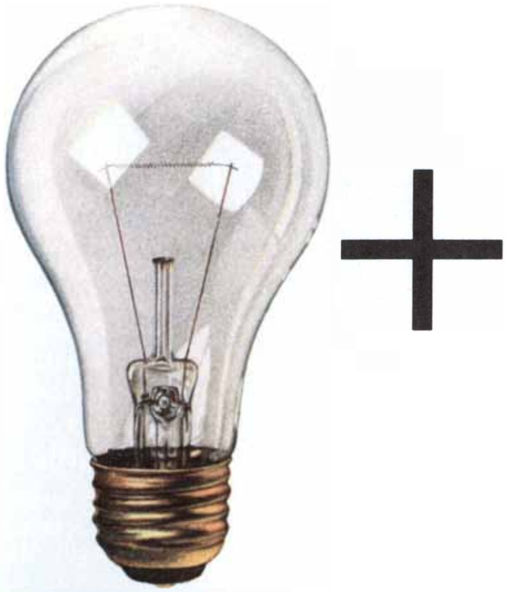
The book is an account of how to live with an animal on terms that are not those of a pet but instead fit a creature whose "ways are those of the wild and must be respected." Phoebe is strongly nocturnal; raccoons that feel safe will forage before dusk and after dawn, but not Phoebe. She stirs first at dusk, comes out of her basket by dark, spends an active evening, drapes herself for a midnight rest on some shelf and grows busy again in the darkest hours before dawn.

But her people, diurnal primates, slept next to her room; she kept them awake during the small hours by racing about, even nibbling their toes. Mutual adaptation was called for. They undertook to reset her biological clock. That required only a few days. The window shade in her room is now pulled down at midday. The sharp cutoff of daylight resembles the rapid nightfall of her tropical forest; she judges that the sunny time of sleep has ended. Phoebe stretches, yawns, grooms and slowly begins her active hours. Sometimes she comes downstairs in the evening, but she soon goes back up to her own dark room with its climbable tree branches and high walkway. At midnight the light comes on in there; that surrogate sunrise signals her bedtime. Tired Phoebe turns in, although it is never too late for her to enjoy a nightcap marshmallow offered by a friend.

In this matter-of-fact illustrated text we read of the simple physiology of Phoebe's low metabolism at rest, of her relish for bananas and grapes enriched by smelts daily and an egg now and again. She uses her scent glands to mark favored people and door tops with distinct musklike odors. She is a keen listener and makes many noises of her own; her eyes are more for shadowy images in the darkness than for detail. Her touch is delicate, even though her somewhat bearlike forepaws are not as deft and long-fingered as a raccoon's.

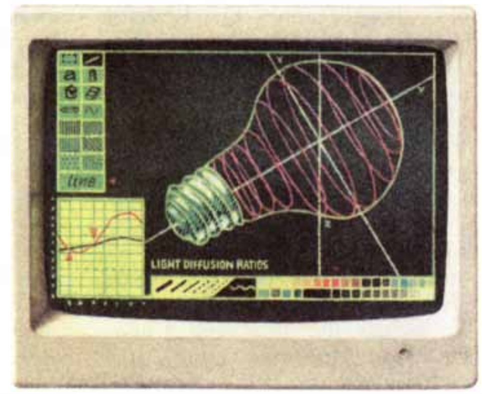
Here is the evidence that close observation, the literature of science and depth of understanding make indispensable human contributions to an unlikely partnership. There was a song of the 1930's that went: "When you do the Kinkajou,/You'll dance before you think you do...." Any grade school reader who has ever seriously wanted to make a pet of some unfamiliar animal will learn from and delight in Phoebe's story. There are many ways to dance the Kinkajou.

BUFFALO: THE AMERICAN BISON TODAY, by Dorothy Hinshaw Patent. Photographs by William Muñoz. Ticknor & Fields ($12.95). SAVING



Who has the tools
to make your idea shine?

Who can listen, and assess the ways information and communication systems can further your idea?

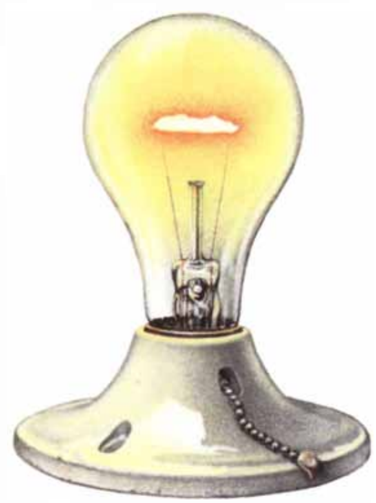


Who has over 5,000 engineers dedicated to providing the best means to enhance your idea?

Who can draw upon technology from around the world?

Who has a proven record of service with a wide range of customers?

Who can offer the dependability and stability of a growing $20 billion corporation?



Who can help make a bright idea a brilliant success?

THE ANSWER IS NYNEX

Telecommunications • Information Systems • Software • Publishing • Business Services

For information, 1-800-535-1535

THE PEREGRINE FALCON, by Caroline Arnold. Photographs by Richard R. Hewett. Carolrhoda Books, Inc. ($12.95). Once the buffalo roamed, 50 million of them, from Reno east to Buffalo (where else?) and from Yellowknife south to Tampico. In 1900 the watchers could count only 20 head in the wild, all in Yellowstone Park. Now the shaggy bands of tens to hundreds of animals of all ages graze on many scattered green mountain meadows and prairie pastures, private and public, across the U.S. and Canada. Unfenced Yellowstone Park still counts about the largest herd, a score of bands; in all perhaps 100,000 bison live wild. "The American bison has a secure future"—if we have.

A new ecology of European cattle, Mesoamerican corn and many exotic Americans has replaced the old symbiosis on the plains that was stable until the horse, the rifle and the railroads came. Brief text and lens look together here at great woolly heads and slender calves season by season: the big bison able to plow away many feet of snow to uncover the grass, the jumpy little calves close to their protective mothers in early summer, bulls fighting in the season of rut as summer ends, and the October roundup on managed range. A list of a couple of dozen public parks where these tokens of the wild past can be observed closes this brief, clear, well-illustrated book for any grade school reader who likes animals or hankers for the legendary West.

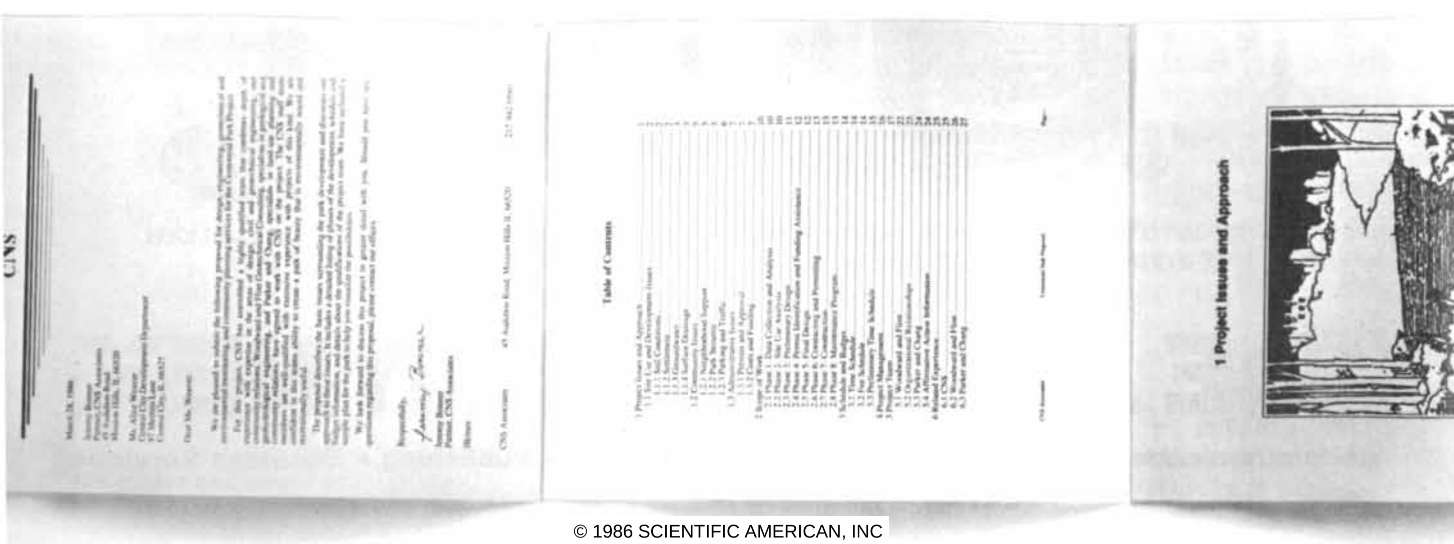
The stoop of the peregrine falcon at dive-bomber speed is newly familiar in the high-rise cliffs of office buildings from Los Angeles to Nairobi, and more frequent in its old haunts too. Its prey are the crowds of pigeons and the small birds of the land. This fierce and beautiful bird was facing extinction a decade and more back, both by loss of habitat and by the effects of pesticides stored in the flesh they fed on. The simple text is made vivid by attractive photographs in color. We see fuzzy chicks and grown birds and devoted people climbing mountain rocks to claim thin-shelled eggs for artificial incubation and hatching in a few centers, perhaps doubling the rate of survival of the young.

Organized effort, including weeks of human help with feeding and nesting, has at least for the present staved off the end of the peregrine. The account (graspable by any moderately skillful reader) centers on modern bird-rearing practice, from the sounds of radio in the breeding cages to helicopter transport back to the remote aerie. This is a tale of knowing intervention, surely a faint echo of falconry. Our species and these raptors preserve their old and paradoxical bond.

Technology and the Physical Sciences

RIVER AND CANAL, written and illustrated by Edward Boyer. Holiday House ($11.95). In only 20 years, between 1815 and 1835, more than 3,000 miles of navigable canals were built in eastern North America. Canals were the easiest and cheapest means of shipping grain from the new farmlands and sending goods in return. Among the lesser-known entrepreneurs of that time were Ezra Horne and Joshua Warren, who formed a company to build a canal from the fall line of the Pocosink, nowhere navigable as it came down the mountains, up to the divide at Nesco Gap on Allegheny Mountain. There would be a short portage to the Bleak River, then the canal would continue westward down the Bleak to the Ohio. The company was chartered by the legislature amidst great enthusiasm in "A.D. 1817, and of Independence the 51st Year."

A fine spread shows a period "aerial view" and a full section of the Pocosink Navigation, drawn up for the company's 30th anniversary. All 280 miles are mapped and profiled; the first locks lie only a few miles up the Pocosink, at Little Falls. Nesco, 200



mule-drawn miles west, is reached after a slow ascent of 726 feet through some 15 sets of multiple locks, 10 feet of rise at each lock. The railroad carried half-size boats over the summit, pulling car and boat by eight-inch hempen cable, powered by the largest steam engines on the continent. The summit was crossed nearly 600 feet up, and then the boats went downslope by rail to the Bleak River, where locks led down to the Ohio. The canalboats were drawn on a marvelous suspension aqueduct from just above the mouth of the Bleak to the final turning basin across the Ohio. Ten-inch cables built up of thousands of wrought-iron wires cleverly spun in place bore the aqueduct's weight. The structure was inaugurated in 1848; the backwoodsmen "were astounded by the sight of a 140-ton canal boat gliding 48 feet above the mighty Ohio."

The author-artist is a landscape architect and a devotee of canals. He would concede that the Pocosink and the Bleak flow only from his pen, but there is real precedent for everything he has invented. The railroad portage was built for canalboats across the Alleghenies in Pennsylvania, and its summit in Cresson is preserved today as a National Historical Site. It was John Roebling himself who built a suspension aqueduct at Lackawaxen, Pa.

The Pocosink Navigation Company survived for only a century, like most of its real counterparts. Its masonry, navvywork and engineering were more enduring than its economics; the remains of the canal can still be found if you go down by the river. The book ends with a list of 20 of the most informative exhibits that mark the remains of real canals in the U.S. and Canada. Some of course are still in use one way or another. The Rideau Waterway runs for 123 miles of navigation, through 47 locks south from Ottawa to Lake Ontario that are today "almost exactly as they were built." Clarity of technical and historical detail in text and picture singles out this rich book; it is easy to read and yet can repay detailed study, in particular for good readers who can hope to visit some lock or quiet canal one summer day.

PLANE CRAZY: A CELEBRATION OF FLYING, by Burton Bernstein. With illustrations by Edward Koren. Ticknor & Fields ($16.95). You can't go home again, some say. But you can, if you are seized with passion for flight, solo "for the first time" three times in one life. The author soloed at 16, elated and trembling in a rattly yellow Piper Cub. He tried again one preoccupied flightless decade later, now an aspirant for a New York writer's career. That time his flying lessons were on the G.I. Bill; he took off in a rattly yellow Aeronca. The third time he soloed he was a middle-aged father, novelist and established writer. Having gone 20 years without flying, he took off in a svelte but intimidating brown-and-white Cessna 152, from an airfield near his Danbury, Conn., studio for what certainly seemed the first exhilarating time—again.

The eloquence of Antoine de St.-Exupéry and the perceptions of Wolf Langewiesche long ago made classic the sights, joys, introspection and fears of the airplane pilot. This straightforward new narrative is equally an account of the joys of flight, although its topic has widened, for flying is now as much a culture as it is an experience of hand and head. A million people fly the airlines every day; it is less the breathtaking views from aloft than the recognition of a life-and-death responsibility, acceptable only through unfailing craftsmanship, that singles out the joyful private pilot from us gawking and irritable passengers.

One coterie we meet is that of the stunt fliers. There is a page or two of history, in which we learn of the very first time the loop was looped, over Kiev in 1913 by Lieutenant Peter Nikolaevich Nesterov. He reported the famous blood-draining blackout; the

Imperial Air Service reprimanded him for so endangering the Czar's property. An account is provided of aerobatics today through the work of Stanley Segalla of Canaan, Conn., whose stuttering, stalling, finally silent maneuvers in a low-flying Piper Super Cub begin as outrageous burlesque and close in a display of utter mastery.

Even more winning is the story of the people who inhabit the small airports, the centers of private flight that dot the American countryside. We visit several, on island or in farming valleys, to share delight and disappointment at these junctions of sport, utility and art. The little community at the lunch counter, the lonely sound of the flag snapping in the wind and the intense and palpable concern felt for every single takeoff and landing—all of these signal this life.

Our ethnographer could not remain aloof; the last of his visits was to an airport at Post Mills, Vt. There one of the owners turned out to be a record-holding glider pilot. Her skill tempted Bernstein to learn anew, now closer to the flight of hawks. The final stylish pen-and-ink drawing is surely intended to entrap the by now persuaded reader (eighth grade and up). It shows a light plane dragging a banner across the upper Connecticut valley: Fly & Soar Post Mills.

THROW IT OUT OF SIGHT! BUILDING AND FLYING A HAND-LAUNCHED GLIDER, by Lawrence F. Abrams. Dillon Press, Inc. ($10.95). Any boy or girl who can manage a pretty good overhand throw and a fair number of hours of purposeful and painstaking handwork can quite literally toss a small glider, wingspan about a foot, out of sight. Lawrence Abrams offers a careful procedure for building such a glider, a design he has flown in 5,000 test flights. There is no explicit theory, although good reasons are given for the detailed instructions. A small group of youngsters who will work together could use this book very well; reading it without actual construction is of less interest.

The chief materials are a few dollars worth of balsa wood, varnish and superglue. The main tools are a sharp knife and lots of sandpaper; of course the specifications are much more precise than these generalities. Here are patterns in full size, long step-by-step instructions and closeup photographs of the process and parts. One essential is the airfoil section of the wing; the form is made by careful sanding to fit the pattern shown. A dime is required as a thickness gauge.

Once the parts are cut, shaped, glued and finished there comes first flight. The plane must be carefully trimmed before a power launch, "a very strong throw upward." Retrimming done, the next step is a timed flight in still air. If you can keep the plane up for 20 seconds, you have a glider that can exploit thermals as the hawks do, and some morning you may watch your own little plane soar visibly for five or 10 minutes, thousands of feet up into the sky, perhaps to disappear into the clouds. It can come back; there is a little spoiler you can make and mount, timed by a slow-burning rope fuse, that will, with luck, send the craft spiraling down out of the thermal and back into sight.

The road to competition in this narrow but remarkably simple form of model aeronautics is briefly charted here too. Serious contestants use much larger gliders. Their record for still-air endurance is 93 seconds aloft inside an airplane hangar. Outside, borne by the thermals, gliders marked with an address have been returned from tens of miles away, after the better part of an hour in flight.

WORKBOATS, written and illustrated by Jan Adkins. Charles Scribner's Sons ($11.95). "I want to write a truthful book about men and women and the sea," without calling on swashbuckling pirates and uncharted islands. "The big sailing ships are gone, and the islands all have names." That is certainly true around the intricacies plotted on chart 131416, representing coastal waters from Provincetown to Block Island. Jan Adkins has drawn that same chart to open his true book. Lying on it are his tools, calculator, compass, pencils, pens and parallel rules.

The tale is about a day when a storm was coming. No one would take out a small motorboat for fun. Only workboats large and small would be in the bay, along with the Coast Guard vessels that safeguard them. Peter Farrel had set out in his little Boston Whaler for some scalloping before the seas rose, but he had not come back after the weather turned.

This is the story of the small search they made for Peter that fall day, and of the people and the boats that took some part.

They were many, from his friend Skip in the yard boat, who was the first to miss Peter, to the big Coast Guard C-130 that flew at low altitude on its cargo run to help look, the car ferry out of Woods Hole, the buoy tender *Bittersweet,* the dragger, the helicopters, the tugs.... We see all of them, drawn with careful line and shade, and more: all the types of buoys, the launches and cutters of the U.S.C.G., all the ferries and the tugs, the floats and nets and rakes and traps of the people who fish in coastal waters.

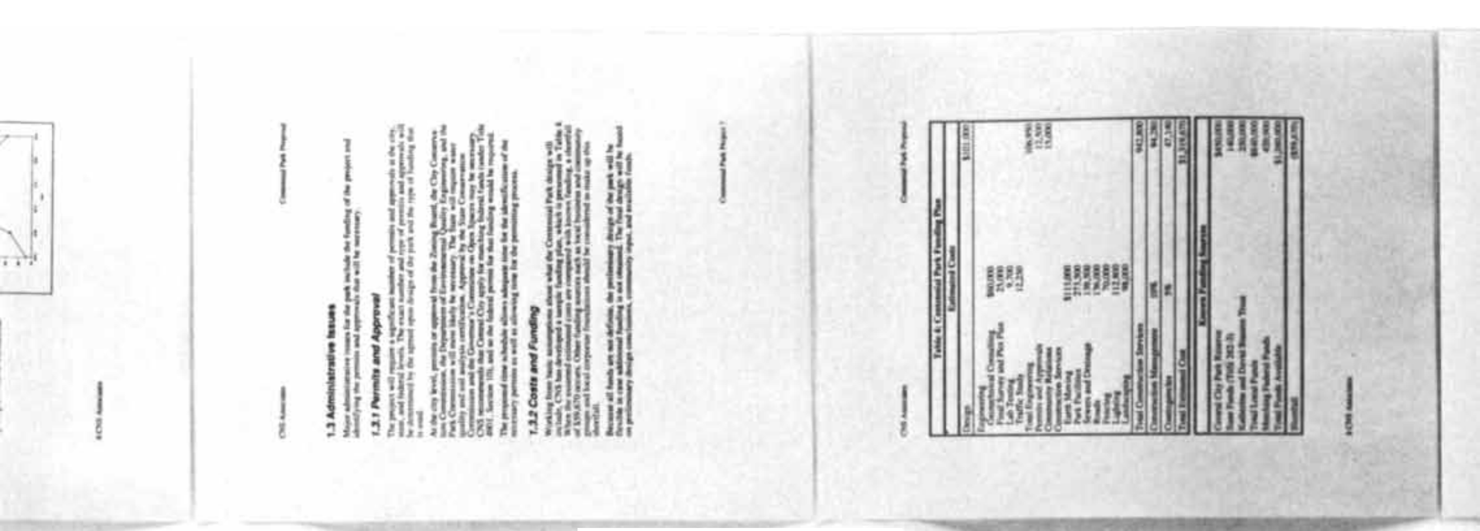
There was a false alarm: an empty boat was seen adrift, but it was not Peter's. A tug captain pushing a barge full of oil south through the canal had listened to the radio alerts all morning; from his high pilothouse he caught sight of something in the marsh as he entered Buzzards Bay. He could not go in there, or even stop for a closer look, but he called up the Coast Guard as he plowed on.

The right knowledgeable man in the right shallow-draft boat picked Peter up before long where he sat waiting on the bow of his beached Whaler. His arm had been hurt when the outboard's propeller had hit a rock; the incoming tide had carried him along. The two came home together, the Whaler under tow.

The text is as full of quiet warmth as are the drawings of well-understood detail. "The most truthful parts of the book are the boats and the common, everyday courage of the people on them." Salty young readers who have a penchant for how boats work will pore over the drawings for a long time: finest kind.

A MODERN DAY YANKEE IN A CONNECTICUT COURT, AND OTHER ESSAYS ON SCIENCE, by Alan Lightman. Viking (paperbound, $14.95). Last year a Bostonian, assistant manager of a department store, found himself lying on the grounds of the Colt Armory in Hartford in the summer of 1880, hurled back in time by a chance blow on the head. A passerby in a buggy stopped to help. "You from New York?" the man asked. (It was the fancy clothes.) They locked him up soon enough, charged with lunacy, once he had tried to talk the Colt machinists (who must have been a little provincial) into starting on an internal-combustion engine. "All you've told us are the names of things," they said, reasonably, as he burbled on about gas and air and spark plugs. Acquitted at his lunacy trial, not by his vague testimony but by a ballpoint pen, the Yankee is sent back to 1985 by another accident, and we are left to muse over counsel's mocking question: "Do you mean to tell me only a handful of people from your century understand how these things work?" Nor was it different then; that Hartford judge could not explain the telegraph.

Alan Lightman is a working theoretical astrophysicist with a flair for informal and easy-to-read prose over an unusual range of stances and forms. In this little volume there are two dozen of his recent magazine pieces, a few as tongue-in-cheek as the tale of the un-



happy Yankee, a few more as personal as the delightful account of the author's undergraduate effort to become an electronics adept. That one-term struggle included covert correspondence-school approval of the design of his project and plenty of hard work and lost sleep. It ended on the appointed day of demonstration in the little disaster of "a single blinding flash of light.... There are abstractionists and there are tinkerers, and I was not unhappy to have discovered my lot."

Not all the essays are lighthearted. The creativity of young scientists is described as "an unwanted glimpse of my mortality," coming unreasonably early in a physicist's career; "private discoveries...are not as frequent now." Topics include what we know of snowflakes today, E. T. calling Harvard, a smile described in elaborate psychophysical detail, as it were photon by photon and cell by cell, a glimpse of Walden Pond and a grimly modest proposal for a demonstrative nuclear war. They form a collection any youthful reader about science in general should enjoy.

Anthropology

THE PUEBLO, by Charlotte and David Yue. Houghton Mifflin Company ($12.95). In a little more than 100 pages of clear prose, sympathetic but not sentimental, the text and drawings of this brief book describe the Pueblo people of today and yesterday as well as many a more ambitious volume has. A section of a mesa, labeled to show the aquifer, evokes the land and the needs of the people. There are a good simple map of the region and a wealth of visual images: the tree rings that date the history of the Southwest, ears of corn, the foundation of Pueblo life; a generalized sketch map of a mesa-top village amidst its orchards and fields; houses of masonry and adobe, and how such houses are built; ladders, ceremonial kivas and dancers, and even a little pile of piki, the paper-thin rolls made from finely ground cornmeal that anticipated breakfast cornflakes.

Emphasis is placed on the material culture of the people, and yet their complex social and inward lives are not neglected. The kachina is here; the sacred cavity in the kiva is described and sketched. Like the Pueblo culture they present, these authors hold that a balance between spiritual and practical purposes is important. The kivas underground are cool in summer and easier to heat by winter; they are also imitations of the world below, where life began. The final drawing shows a pickup truck parked by the new adobe house; most Pueblos live now in the world, closer to fields and springs, but they return to the old, defensible village for spiritual community and the sense of traditional values. "A Hopi bridegroom still receives his gift of corn meal on a handmade basket tray." There is a good bibliography that leads to detailed information. Accessible to readers in the middle grades, the book would serve many an adult reader as a serious introduction to one of the most enduring human creations on American soil.

THE ENCHANTED CARIBOU, by Elizabeth Cleaver. Atheneum ($9.95). The Inuit hunters by summer pursue the caribou that roam the tundra. The people live in summer tents of caribou hide and dress in the skins of caribou. They know magical songs and dances that ensure success, and they perform them at need. They never take the rare white caribou; instead they "treat it kindly and do not kill it, for it might be enchanted."

This little tale of the Inuit recounts the events that led to the enchantment. It is a story of love and magic, here retold simply, poetically and rather personally by a Montreal artist who has written and illustrated a number of poems and tales the world around. It came to her that this story, which is based on visions, dreams and the magical transformation of a young woman into an animal, might be well suited to the shadow-puppet theater, whose technique she had once taught to Inuit children at Baker Lake.

In this book each page of simple text is faced by one of picture. The scene was made from a cutout in black paper, strongly lighted from behind and photographed in black and white through a translucent screen. A page of helpful instructions and a couple of patterns for the figures are given at the end to help any readers who care to make a shadow-theater version of the story for themselves; such a version can of course move, as no book can. This is a book young people can both read and work from. We sorely miss any statement of where Elizabeth Cleaver found this story of the enchanted caribou.

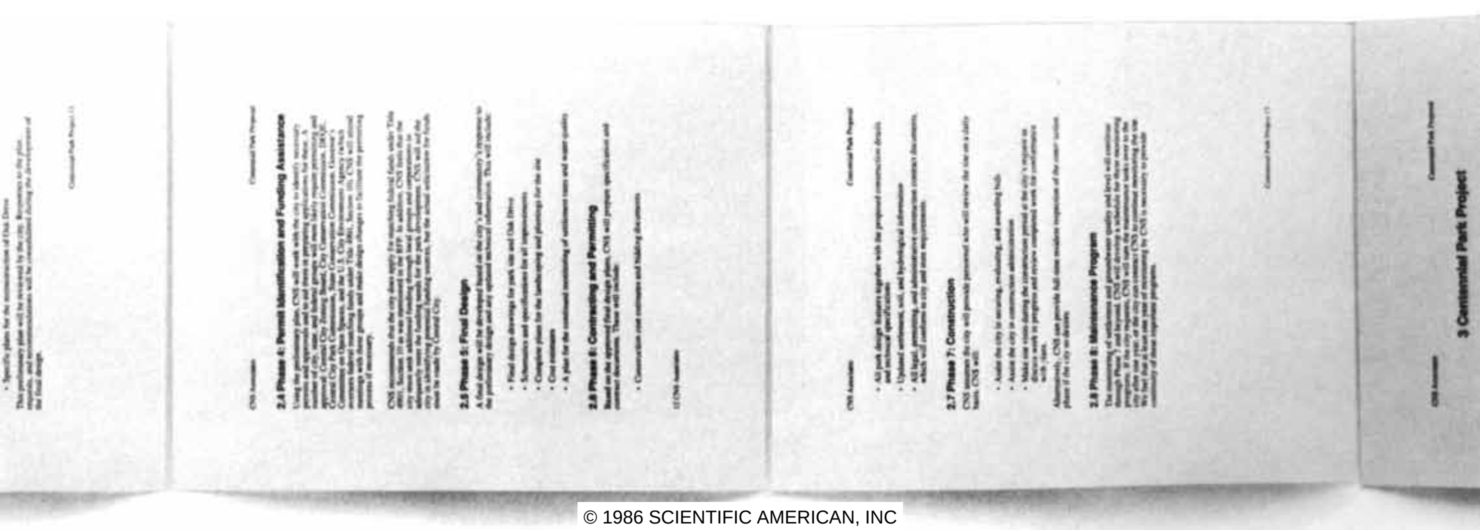
THE WEAVING OF A DREAM: A CHINESE FOLKTALE, retold and illustrated by Marilee Heyer. Viking Kestrel ($12.95). Eighteen or 20 large, colorful painted pages, more crowded than any remembered dream, give this thin volume its unquestioned distinction. They display fantasy mixed with a thoroughly Western representation of the landscapes and costumes of old China; a bat grimaces, big against the moon; below the two in the mist stand the strange limestone pinnacles along the River Li. The images are all rather cinematic; the drama is told in a storyteller's manner, full of ordeal and betrayal, on the facing pages of text. Carp, serpents, lizards, butterflies, all foreground-large, compete for attention with the brocades woven by the poor widow whose talent was so great that the fairies wanted only to copy them. There is a happy ending for all save the worst of the people in the tale.

Mathematics and Perception

PUZZLES OLD & NEW: HOW TO MAKE AND SOLVE THEM, created by Jerry Slocum and Jack Botermans; text by Carla van Splunteren and Tony Burrett. University of Washington Press ($19.95). Martin Gardner himself happily reports that this book presents, "for the first time in history,...a comprehensive survey" of mechanical puzzles. The term is used for those puzzles made of solid parts that must be manipulated to arrive at a solution. Even a count of the objects displayed here is difficult, but there must be between 400 and 500 distinct puzzles. This volume arises out of the annual international puzzle parties that are held around the prodigious collection of Jerry Slocum of Los Angeles, and it enhances a traveling exhibition originated by the Craft and Folk Art Museum of Los Angeles.

The authors have an organizing taxonomy. They distinguish 10 classes of these puzzles: take apart, put together, interlocking, disentangling, sequential move, vanishing image, object impossible to make, folding, enigmatic vessel and puzzles demanding manual dexterity. One essential page shows an array of 100 or so realizations of Rubik's (or Larry Nichols') Cube, a world teaser a few years back. Another set of pages cracks open a hard chestnut of enduring influence, the Chinese Rings. That wiry disentanglement requires movement sequences—which are mapped out here—that recall the equally classical Tower of Hanoi. No taxonomy rests easy.

The variety of puzzles represented can be described only as awesome. In contrast to that notorious cube, another spread shows four or five dissections of plane figures. An outline letter T, for instance, is divided by two parallel cuts into four parts. Of unsurpassed simplicity, this cheap T puzzle is a kind of logical converse of the commonplace jigsaw. Its origin goes back to the turn of the century; it has periodically become popular. In Chicago it was circulated (to advertise sausage) as the Tormentor, the handbill proclaiming



that it was "a real test" to fit the parts back into a perfect T. The earliest version on record was a five-part ivory cross, made in China, widely known 150 years ago.

Most puzzles evolve. The interlocking wood puzzles of Edo Japan surely first reflected the skill of carpenters who fitted wood pieces without nails. Those puzzles are made today in architectural and animal forms of all kinds, as well as in the shape of cars, ships and planes. Puzzles like those but more abstract, interlocking notched rods assembled into "burrs" that show high external symmetry, have been made for a long time in the West; there are said to be some that consist of hundreds of pieces.

A unique specimen is a bolt and nut a foot high, a 1972 puzzle patent. The nut can be unscrewed from the bolt threads, a series of spaced annular grooves, only by a specific sequence of turns; the puzzle is a rotary maze! It is foolhardy to summarize more of this cornucopia of bizarre ingenuity. The book will transfix some readers at almost any age; others, chilled, will put it down at once. One wonders, though, whether any library that serves a population of teen-agers can afford to be without a copy. Is its topic mathematics, perception, artisanry, recreation or a benign voluntary insanity?

OPTICAL ILLUSIONS, by Laurence B. White, Jr., and Ray Broekel. Franklin Watts ($9.40). Adroitly simplified without condescension in language and scope, this book by two experienced authors who have written on magic for young people quietly opens a gate to scientific inference. The examples include tried-and-true length misjudgments (given a new look), the blind spot, colored afterimages, a flip book and similar classics. The reader is urged to look in a mirror at three bold-letter texts here, with the book held upside down. The texts are artfully constructed. The last one says KID—/DO HIDE/BOX & BIKE (the ampersand is here drawn to be symmetrical up and down).

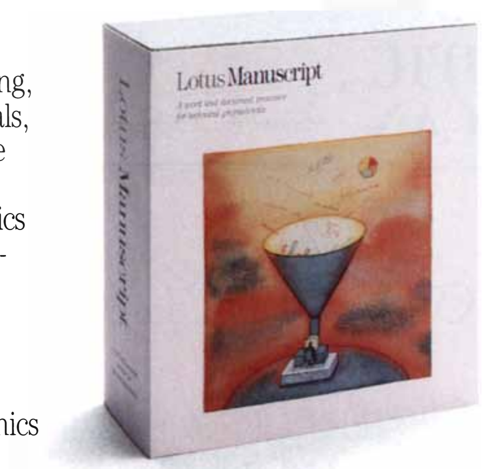
The light touch and fresh pedagogy make the book an unusually attractive stroll along a familiar path. At one point the text argues from what we have learned in favor of the value of such measuring tools as rulers; the next step, that inferences from instruments can fail as well, is not made explicit. Such epistemological optimism is justified for young readers.

The visiting card of magician Orville Meyer presents a dazzling example of distorted perspective, that "long, skinny writing" legible only at near-grazing incidence. The four-color process is effectively displayed by reproducing the intermediate steps in printing a watercolor rainbow. A final chapter puts some of these notions to use in the simple theater of children's magic. This is a first-rate start for fourth-grade readers and up.

SHAPES, SHAPES, SHAPES, by Tana Hoban. Greenwillow Books ($11.75). Once more we look through Tana Hoban's discerning lens, in another book she has made out of two dozen handsome photographs without words (an initial list of shapes is offered). This time she is on the lookout for simple geometric forms, for circles, triangles, stars, ovals, arcs.... She picks them up everywhere, uniting the world. The child who has playfully set some half-dozen bright plastic barrettes into her hair joins with the tourist boat that glides along the Seine, carrying high its own gleaming array of white life-preserver rings. The exposed wall of a house that has somehow lost its neighbor bears marks of all the vanished floors, walls and stairs, the geometry of the architectural elevation become full size. Pylons and ladders and cookies and hopscotch are only a little of what is here, giving pleasure as so often before, attractive for prereaders (and post-) and yet a serious first lesson in the subtle affinity between the real and the abstract.

Yes, Lotus Manuscript™ is a very different kind of word processor. Because when you're processing long, technical documents such as proposals, specifications and reports, you have very different kinds of needs.

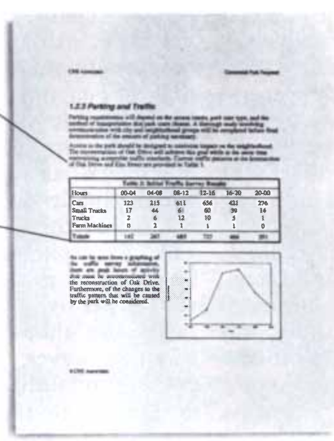
Needs like mixing text and graphics on the same page. Or, creating complex equations. Or, working with columns and tables. Or, making wholesale revisions.



Very often technical documents require the mixing of text and graphics on the same page. Graphics from Lotus® 1-2-3,® Symphony® and Freelance® Plus. With Manuscript you can import diagrams, and scanned images, as well as spreadsheets and charts.

Another feature designed specifically for the writer of technical documents is Manuscript's integrated Outliner. When you collapse the document you're able to navigate the outline much easier than wading through the entire piece. You can move a page, or even an entire chapter with just a few keystrokes.

Also, because Manuscript is structured, you can globally format an entire document or format by individual sections. Imagine changing all headlines from 14 point plain text to 16 point bold, all with just one command.



Manuscript allows you to mix text and graphics on the same page.

You can throw away your scissors and glue, cut and paste are a thing of the past.

How about the age-old dilemma of proof reading revisions? Manuscript's Document Compare feature highlights changes between revisions. You might call it fool-proofreading.

Also, Manuscript automatically sizes and generates math equations as well as upper and lower case Greek symbols, diacritical marks and brackets. No longer do you need to literally cut and paste to create complex equations.

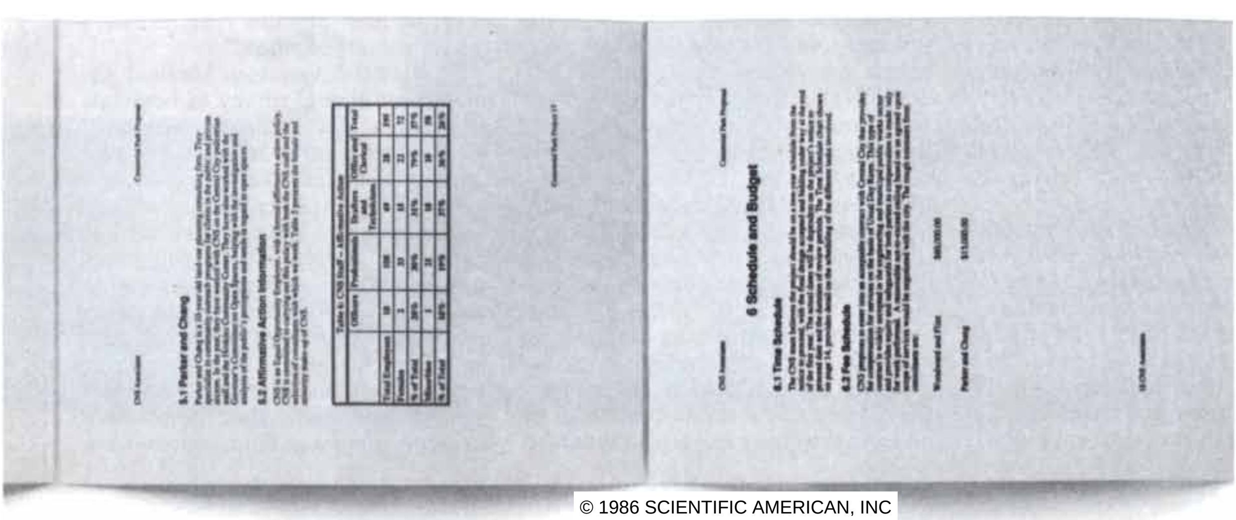
With our powerful Print Formatter you can control the look of your document with the creation and editing of tables. You can choose fonts and point sizes from a built-in typeface list. You can even save format information as a template for regular use.

As for the quality of output, Manuscript takes full advantage of today's printing technology, from dot-matrix to laser, including PostScript® printers.

With our Document Preview feature you can see exactly how any page will look before it's printed.

Manuscript is designed to work on most IBM PCs and compatibles. Its familiar 1-2-3 interface makes it easy to use. And our Manuscript evaluation kit makes it easy to try. For $10.00, you'll get a presentation disk, working software, and a tutorial manual. To get your evaluation kit, call 1-800-345-1043, ask for lot #BD-1450. Or, for more information, see your authorized Lotus dealer, or write Lotus Development Corp., 55 Cambridge Parkway, Cambridge, MA 02142.

May we suggest Lotus Manuscript.



Established 1845 SCIENTIFIC AMERICAN December 1986 Volume 255 Number 6

Corporate Medicine for Profit

For-profit chains of hospitals emerged from a long tradition of commercialism in the U.S. health-care system. They have thrived on the system's excesses and highlighted its inequities

by Donald W. Light

A new era in American health care can be said to date from 1968. In that year Thomas F. Frist and Jack C. Massey—the former a Nashville doctor and the latter the man who had made Kentucky Fried Chicken into a national chain—formed the Hospital Corporation of America (HCA) to provide Frist's private hospital, Park View, with capital for expansion. HCA soon began acquiring additional hospitals, and it is now the country's largest investor-owned hospital chain. At the time of its formation few of the nation's non-Federal, acute-care hospitals belonged to a profit-making chain. By 1983 for-profit chains controlled 13 percent of those hospitals.

The transformation brought about by the for-profit chains is more extensive than the percentage suggests. By force of example and direct competition the for-profit chains have driven many nonprofit hospitals also to combine into chains. Today about a third of the country's acute-care hospitals belong to multi-unit systems. Investor-owned corporations have also established themselves in many other areas of health care, ranging from primary-care clinics to specialized referral centers. It is in the guise of for-profit chains, however, that the corporate presence in health care provokes the most debate.

The juxtaposition of the commercial ethos familiar in fast-food chains with hospital care challenges traditional images of medicine as the embodiment of humane service and even charity. The investor-owned chains have elicited a number of specific criticisms as well, voiced most powerfully by figures in academic medicine such as Arnold S. Relman, professor of medicine at the Harvard Medical School and editor of the *New England Journal of Medicine.* He and others believe commercial considerations could undermine the responsibility of doctors toward their patients, conceivably leading to unnecessary tests and procedures or—given other financial incentives—to inadequate treatment. Critics of for-profit chains also suspect they drive up the cost of health care, reduce its quality, neglect teaching and research and reject those who cannot pay for treatment. As the Federal Government and private insurance companies attempt to limit spending on health care, cutting into hospital profits and forcing hospitals to reduce their own costs, many of those concerns have intensified.

Research that addresses criticisms of the for-profit chains is beginning to accumulate, notably in the form of a comprehensive study recently completed by the Institute of Medicine of the National Academy of Sciences. Considered together with the institutional and economic context from which the chains emerged, the evidence makes it possible to assess their role. Some of the charges that have been leveled at them miss the mark but others are confirmed. The failings of the for-profit hospitals are not theirs alone, however. Those hospitals are only the purest expression of a commercialism that has come to pervade American medicine. They have unabashedly exploited a system that is devoted to the best possible care at any price for most of society but is effectively blind to the needs of the rest. In doing so they have served as a sensitive diagnostic of the system's inequities and excesses.

Only the scale of the commercialism embodied in the for-profit hospital chains is new to American medicine. Although most hospitals have traditionally been "voluntary," that is, run on a nonprofit basis by local community associations or religious orders, a substantial number have been run for profit by proprietors, usually doctors. In 1928, 38.9 percent of the 4,367 nongovernment general hospitals were proprietary—a much higher percentage than today, in spite of the rise of for-profit chains.

Because they tended to be quite small, however, these early proprietary hospitals had only about 16 percent of nongovernment hospital beds. Doctors often founded them because there was no voluntary hospital nearby or because they could not gain admitting privileges at larger, more prestigious hospitals. The quality of care in such private hospitals was generally poor; patients in some regions called them "buckets of blood."

By 1934 the American Medical Association's annual survey of hospitals included a new category, "corporations unrestricted as to profit," in recognition of a class of hospitals owned by stockholders rather than by individuals or partners. The category encompassed 32.4 percent of the proprietary hospitals in that year and, because the corporate hospitals tended to be larger than other proprietary hospitals, 52.3 percent of the proprietary beds. Over the next three decades some of the early corporate and other proprietary hospitals went bankrupt, some were converted into community hospitals and the rest continued to operate on the fringes of the health-care system.

The roots of today's commercialism reach deep even in the main tradition of American medicine. Between 1870 and 1910 advances in major surgery and in the control of infection led to the construction of thousands of new hospitals. Earlier hospitals had served largely as a refuge for the indigent and the dying. The new hospitals attracted paying patients and introduced principles of business management into their operations. Even charity hospitals, which traditionally did not charge for care and depended largely on philanthropy, sought paying patients as the depression of the 1890's reduced donations at the same time as high inflation increased costs.

David K. Rosner, a noted medical historian at Baruch College of the City University of New York, has found that the trustees of the charity hospitals began to woo doctors who had private practices in an effort to attract their well-to-do patients. Such individuals had previously shunned hospitals in favor of home treatment. At the same time the charity hospitals refurbished their rooms and advertised their amenities: brass bedsteads, open fireplaces, serving rooms, private-duty nurses and a good chef. In 1899 the *New York Times* quoted a young woman as saying: "One is not quite in the swim nowadays if one has not had an operation performed.... A hospital has nothing but pleasant associations for me, and you will never find me out of one again when I am ill." The influx of paying patients not only kept the charity hospitals solvent but also enabled them to continue caring for the poor. Before the turn of the century, as a result, the practice of supporting charity care through higher charges to paying patients was established.

In the process, however, the trustees of the charity hospitals ceded a measure of control to private physicians, who altered the hospitals' traditional mission. Many of the doctors, according to Rosner, were more concerned with making the hospital an up-to-date

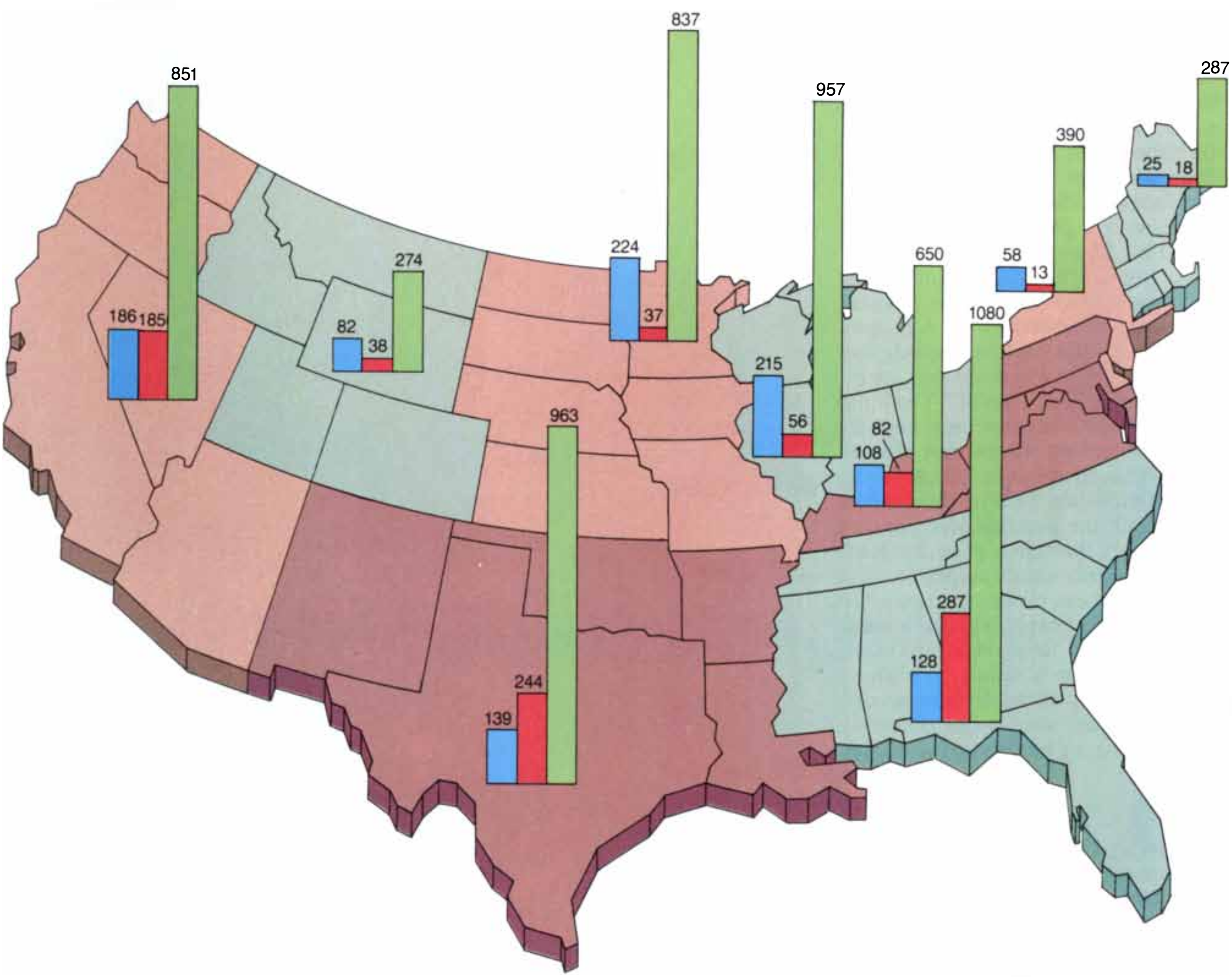
"MEDICAL CAMPUS" in Delray Beach, Fla., offers health care with a corporate face. The complex, owned by National Medical Enterprises, Inc., a for-profit hospital system, combines a general hospital with facilities for psychiatric care, drug-abuse and alcoholism treatment, rehabilitation and convalescence on 40 acres of land. Having prospered as chains of general hospitals, many for-profit systems are now "vertically integrating" their business and offering a range of medical services.

physicians' workshop in which to treat their paying patients than they were with providing charity. Rosner concludes that such hospitals were transformed from neighborhood institutions dedicated to treating everyone into large enterprises offering specialized, up-to-the-minute care and advertising their amenities to a citywide market of affluent patients.

Thus a tradition of running hospitals like businesses evolved early, and the incentives for commercial behavior have only grown since then. As insurance developed first for hospital bills and later for doctors' bills, the hospital industry and the medical profession insisted on reimbursement for the "usual and customary" charges, charges the doctors and hospitals themselves were to have a major say in determining. Whereas European countries legislated fee schedules or established fixed national budgets for health, in the U.S. the growing amount of private insurance and the open-ended reimbursements created a golden era, starting after World War II, during which hospital budgets and physicians' incomes increased sharply.

Soon public funds supplemented the abundance of private-insurance money available for health care. In 1965 Congress established Medicare and Medicaid to pay some of the health-care costs of the elderly and the "deserving" poor respectively. By giving the elderly and some of the poor much more purchasing power than before, the programs significantly increased the number of paying patients. The legislation establishing the programs also included special provisions encouraging proprietary hospitals to expand. It allowed for-profit hospitals to depreciate their capital investment and specified that Medicare and Medicaid payments should not only cover the cost of treatment but also reimburse proprietary hospitals for interest on debt and give them a generous return on equity.

By the mid-1960's, then, physicians, hospitals, private insurers and the Government had devised a system of health-care financing that was ripe for big business. The form it took—the investor-owned hospital chain—reflects a further condition: the chronic need of hospitals for capital to expand, renovate or equip their facilities. Multihospital corporations, unlike privately owned independent hospitals, can raise capital by issuing stock. Moreover, their ability to use packages of hospitals within the chain as collateral enhances their ability to borrow money. Multihospital systems have consid-



MAP OF U.S. HOSPITALS compares the number of short-term hospitals in investor-owned systems (*red*) with the number belonging to nonprofit chains (*blue*) and the total number of short-term hospitals (*green*) in each of nine regions. The proportion of hospitals included in for-profit chains is largest in the South and West, where the population is growing and there are fewer regulations and competing hospitals. For-profit chains account for a relatively small fraction of all hospitals, but they have displayed innovative commercial behavior and stirred controversy. The map is based on 1984 figures from the American Hospital Association.

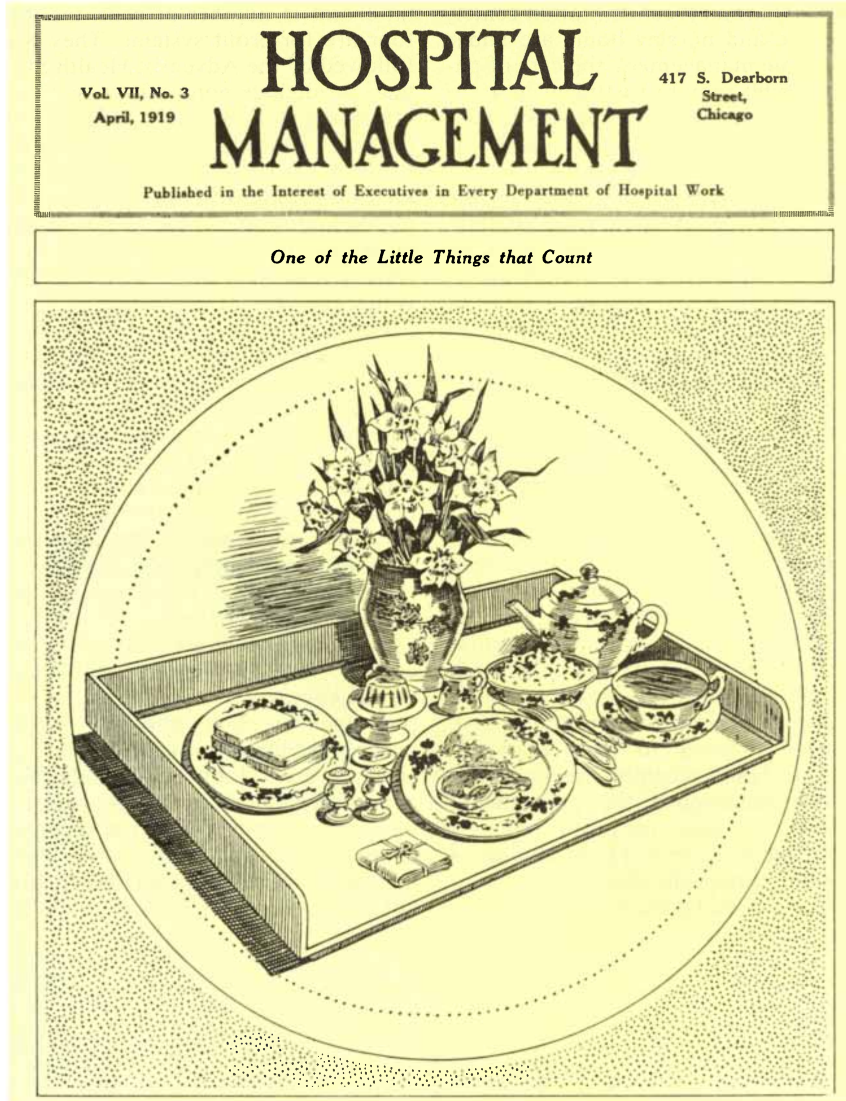
erably better credit ratings than independent hospitals and can borrow at lower interest rates.

During their first years the investor-owned chains grew mainly by buying other for-profit hospitals; in the 1970's more than 80 percent of the chains' growth took this form. At first many of the acquisitions were independent proprietary hospitals, but increasingly the dominant chains bought hospitals from other chains. In the 1980's public hospitals and private voluntary hospitals also began to be numbered among the acquisitions. The managers or trustees of many such hospitals, unable to raise enough capital to renovate aging facilities, have actively sought to be acquired. The chains have also built new hospitals; new construction accounted for 20 percent of their growth in the 1970's but only 4 percent from 1980 through 1984.

The chains have staged most of their growth in areas where the market is attractive: states with increasing population, rising per capita income, few regulations and liberal insurance coverage. They have also sought out regions where there is limited competition from other kinds of hospitals. The chains have found those conditions most often in the South, the Southwest and the West.

Each year from 1977 through 1984 the number of hospitals affiliated with investor-owned chains increased by an average of about 11 percent, and by 1985 the 53 investor-owned chains owned or leased more than 680 hospitals with a total of more than 100,000 beds. HCA, born of the collaboration between Frist and Massey, accounted for half of those beds. Many corporations had also undertaken to employ their financing, management and marketing skills in operating hospitals they did not own: a total of 380 institutions with more than 48,000 beds.

In the past two years the phenomenal expansion of the hospital chains has slowed and their profits have declined. About half of the beds owned by the for-profit chains are now empty in part because of the cost-cutting measures recently adopted by Medicare, Medicaid and private insurers. Hospitals are now, for example, paid a preset fee for many kinds of treatment, an arrangement that encourages them to discharge patients quickly. The oversupply of beds also reflects the overbuilding and refurbishing encouraged by various provisions of Medicare and Medicaid in the previous era of generous reimbursements. The chains are now scrambling to sell off unprofitable hospitals.



ATTRACTING PAYING PATIENTS began to concern large nonprofit hospitals in the early years of this century; the frontispiece from a 1919 issue of a journal for hospital administrators reflects the new attitude. In the 1800's such hospitals had generally provided charity care and offered few amenities. By the turn of the century advances in surgery and in the control of infection drew patients able to pay for treatment; at the same time changing economic conditions led hospitals to depend on fees from the paying patients. The drawing is reproduced by courtesy of the New York Academy of Medicine.

They are also trying to circumvent cost-containment measures, which so far have focused mostly on hospital payments, by diversifying into other areas of health care. Psychiatric hospitals have been a favored area; insurers have not yet placed limits on their reimbursement for psychiatric care. The demand for beds in private psychiatric hospitals, moreover, is expected to rise as government-run hospitals cut back their services or close and as the stigma of psychiatric care diminishes. Nursing homes, which have traditionally been run for profit by private owners, have proved attractive to the hospital corporations as well. The corporations have also established themselves in many other kinds of health-care service: surgicenters (for ambulatory surgery), renal-dialysis programs, primary-care clinics, alcoholism and drug-abuse centers, hospices, industrial medical centers (for sick or injured workers), health-promotion programs, PPO's (preferred-provider organizations, which organize networks of physicians willing to provide discount care) and HMO's (health-maintenance organizations, which give comprehensive care as it is needed in return for a regular prepaid fee).

The hospital corporations have expanded into other services not only in an attempt to outflank cost containment but also to enable them to sell comprehensive packages of services to cost-conscious buyers. When the hospital, pharmacy, rehabilitation center, hospice and nursing home are under the same management, the care of patients who require treatment in a variety of settings can be coordinated more easily and, in theory, more efficiently. The same coordination, of course, could serve in exploiting the patient or the insurer in order to maximize profits. Some chains are integrating their services even further: they are buying or joining forces with insurance companies, enabling them both to sell insurance policies and to care for the policyholders.

The end of the era of dramatic expansion for the profit-making chains offers an opportunity to judge their impact on the health-care system. Their profoundest effect has been an indirect one: the for-profit chains have led the way to the commercialization of the entire system. Driven by many of the identical considerations that spurred the growth of for-profit chains, in particular a need for capital, clusters of nonprofit hospitals have banded together to form nonprofit chains. Until recently the nonprofit chains outstripped the phenomenal growth of their for-profit counterparts: there are now at least four times as many nonprofit chains, with twice as many beds. Of the 10 largest hospital chains in number of beds, the top four are for-profit systems. They are followed by the Adventist Health System, a religious nonprofit chain. The remaining five include two nonprofit religious systems, a for-profit chain, the New York City Health and Hospital Corporation (a public urban conglomerate) and Kaiser Foundation Hospitals, Inc. (a nonprofit HMO that owns a large system of hospitals, mostly in California).

Nonprofit systems resemble the for-profit corporations in many respects. They start or acquire PPO's, HMO's, alcohol- and drug-rehabilitation centers and the like. They manage other hospitals, invest in real estate and buy supplies at a discount. And they have been making money. Between 1983 and 1984, for example, the surpluses of the secular nonprofit chains increased faster than the profits of the investor-owned chains.

Fundamental differences remain, of course. Instead of paying dividends to stockholders, the nonprofit chains plow their surpluses into new or improved facilities, capital reserves or charity care. Many of the religious chains have one or two money-losing inner-city hospitals that are supported by more prosperous hospitals. Furthermore, most nonprofit chains are not intent on growing indefinitely. In most cases they exist primarily to serve their member hospitals rather than to establish a comprehensive system for the country or a large region. As a result the growth of the nonprofit chains has slowed down considerably. Meanwhile the big investor-owned corporations, even though they are not adding hospitals as fast as they once were, continue to expand into other areas of health care.

In another sense the distinction between the two kinds of hospitals is dissolving. A corporate ethos originating among the investor-owned chains has spread throughout the health-care system. In the past hospital administrators usually did not even know the true costs of the operating room or the intensive-care unit. Now the for-profit chains have set an example for all hospitals by introducing sophisticated management techniques and information systems for controlling the flow of supplies, personnel, patients and dollars. They have adopted aggressive marketing techniques meant to attract well-to-do patients, such as business executives, and patients for whom insurance compensation is generous, such as psychiatric cases and those covered by workmen's-compensation policies. Increasingly, other hospitals are following suit.

Amenities such as streamlined billing and convenient hours have also figured in the chains' marketing strategy, as has an effort to create an image of efficiency and polish. Humana, Inc., for example, is known for specifying the number of seconds allowed for a meal to travel from the kitchen to the bedside and for a nurse to arrive after a patient presses the call button. Now that almost a third of the hospital beds in the U.S. are empty because of overbuilding and efforts by insurers and the Government to cut costs, all hospitals are learning these lessons in commercialism.

The for-profit chains contend that their hospitals are run more efficiently than nonprofit hospitals and offer better value for the health dollar. Most of the 13 studies that have evaluated the claim so far do not support it. In the most prominent study J. Michael Watt, Robert A. Derzon and James S. Hahn of Lewin and Associates, Inc., a consulting firm, and investigators from Johns Hopkins University matched 80 investor-owned hospitals with nonprofit hospitals having a comparable location, scale of operation and mixture of services and cases. In 1978 and 1980, the years for which the workers compiled figures, the investor-owned hospitals incurred higher costs per patient per day in ev-

FOR-PROFIT HOSPITALS made up a large percentage of nongovernment hospitals (*gray*) and all hospitals (*black*) in the early years of this century. (The broken part of each curve represents a period for which data are scarce.) Because such proprietary hospitals were generally small, they controlled a relatively small proportion of total beds in nongovernment hospitals (*light color*) and in all hospitals (*dark color*). Most early proprietary hospitals were freestanding institutions owned by one or several doctors, and many offered substandard care. The recent increase in the number of for-profit hospitals after decades of decline reflects a new development: the emergence of chains of proprietary hospitals, owned by investors. Those hospitals are often of better quality than the older proprietary hospitals, many of which were bought and refurbished by the growing chains.

ery category examined but one: the cost of operating their facility. The differential was largest, ironically, in administrative overhead, for which the for-profit hospitals spent an average of 57 percent more than their nonprofit counterparts. Of that difference, 83 percent reflected spending on the corporate headquarters of the chains.

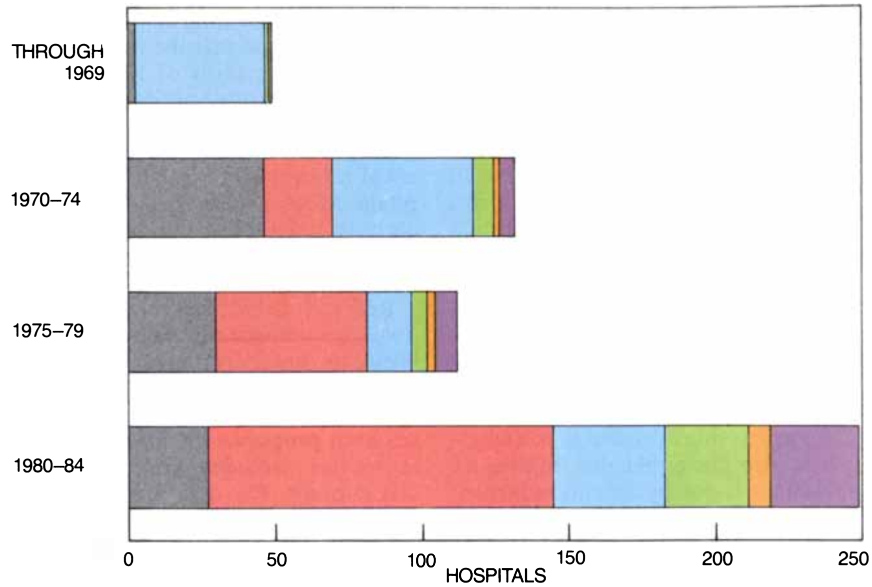
At the same time, the for-profit hospitals studied also charged significantly more than the nonprofit hospitals did: an average of 22 percent more per admission. Routine bed charges did not differ much between the two types of hospitals; the disparity results from much higher charges by the investor-owned hospitals for ancillary services such as tests and drugs. After taking into account the taxes to which for-profit hospitals are subject and the gifts and subsidies nonprofit hospitals receive, the investigators found that the net revenues of the investor-owned hospitals were 10 percent higher per bed per day—enough to make such hospitals more profitable than the nonprofit ones, in spite of higher costs.

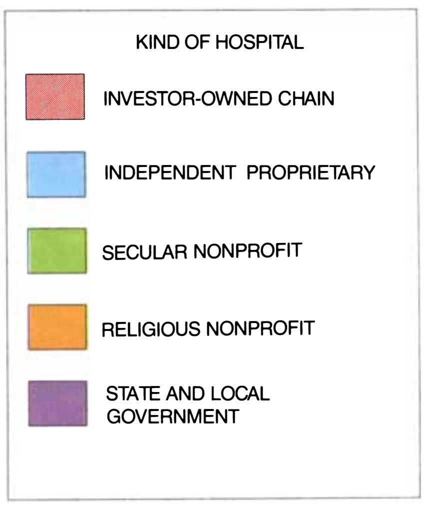
Traditionally, of course, the system of health-care financing in the U.S. has not encouraged efficiency, because it has compensated hospitals on the basis of their costs or their billed charges. Indeed, the system effectively discouraged efficiency: lower costs or charges led to lower reimbursements and made it difficult to justify subsequent increases. The efficient hospital in effect tightened its own financial noose. Now that cost-containment measures are putting an end to open-ended reimbursement, the for-profit corporations will need to achieve the operating efficiencies that sophisticated management should make possible.

Given the lack of incentives to cut costs, it is easy to see why the repeated charges that the chains skimp on quality have not been confirmed in half a dozen research studies that have looked into the question. The for-profit hospitals in the Watt study, for example, employed 13 percent fewer staff per adjusted patient-day than the nonprofit hospitals but paid them an average of 10 percent more in salary and benefits. Conceivably the higher pay attracts staff whose better quality offsets their proportionately smaller number. Other studies have examined such measures of quality as death rates and courses of illness, malpractice liability, hospital accreditation and the proportion of board-certified specialists on the staff at investor-owned and nonprofit hospitals. Small differences emerged in individual measures, but overall the two kinds of hospitals were found to be comparable.

Because quality is an elusive attribute, meaning different things to different people, not all concerns have been dispelled. Do the for-profit chains cut corners, for example, when their hospital is the only one in town or in the county? There is no evidence that they do, but speculation continues. Another concern focuses on the bonuses an enterprising hospital might offer its staff physicians as an incentive for keeping costs down or increasing revenues. No decline in the quality of care has yet been attributed to such an arrangement, but the Institute of Medicine has declared that such bonus plans threaten doctors' obligation to act in their patients' best interests.

The many instances in which a for-profit chain has bought a small, financially troubled hospital provide another perspective on quality. In some cases the corporation has kept the hospital from going out of business, and in many more cases it has spent heavily on renovating and upgrading the facility. The generous provisions of Medicare and Medicaid regarding interest on debt, return on equity and depreciation have allowed much of the cost to be transferred to the taxpaying public. The open-ended reimbursement system has covered other costs and enabled the corporation to hire a capable staff and improve the hospital management. A study by Stephen M. Shortell of Northwestern University has shown that the corporation often gives doctors a more active role in running the hospital than they had before. Thus transformed, the hospital can attract a better medical staff, and quality probably improves.





SOURCES OF NEW HOSPITALS added to investor-owned chains have varied with time. The bars combine statistics for six large for-profit chains: Hospital Corporation of America; Humana, Inc.; American Medical International, Inc.; National Medical Enterprises, Inc.; Charter Medical Corporation, and Republic Health Corporation. The chains at first grew almost entirely by buying or, in a few cases, leasing hospitals (*color*), mostly independent proprietary hospitals. More recently they have also acquired hospitals from other chains, as well as an increasing number of nonprofit and government hospitals. The chains have also grown by building new hospitals (*gray*). Since 1984 an oversupply of hospital beds has developed and all growth has slowed. The graph is based on work done by Elizabeth W. Hoy and Bradford H. Gray of the Institute of Medicine.

Although any hospital exists mainly to treat patients, many hospitals traditionally have accommodated teaching and research as well. Some critics believe hospital corporations give low priority to most teaching and research and may change the character of what they do maintain. Until recently concern about the chains' attitudes toward teaching and research was largely moot, since they initially bought smaller, less distinguished hospitals, most of which had few such programs in the first place.

The chains do consider certain kinds of teaching to be good for business. In the hospitals they have taken over, many corporations have established or maintained training programs for nurses, laboratory technicians, respiratory therapists and other skilled personnel. Hospitals with good educational programs offer opportunities for advancement and hence attract higher-caliber personnel, and a better staff attracts more profitable patients.

It is also clear that the training of physicians is not by definition incompatible with corporate goals. HCA, for example, now owns or manages 34 hospitals with residency programs; 64 of HCA's hospitals have medical-school affiliations. After extensively publicized negotiations several large teaching hospitals recently became allied with other investor-owned chains. In such arrangements the university is often motivated by a lack of capital needed for renovating an aging facility and by the hope that the management and marketing skills of the corporation will render the hospital financially sound. The corporation, for its part, gains prestige, visibility and a sophisticated referral center for its other hospitals. Most observers agree it is too early to judge how corporate control will affect the character and mission of teaching hospitals.

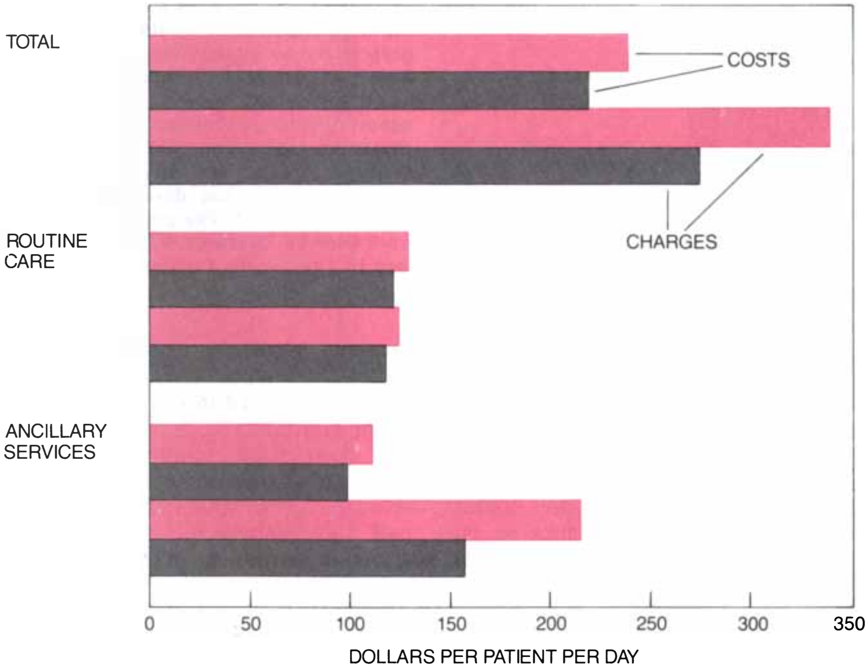
Research as well as the training of doctors is a major function of teaching hospitals, and corporations that have taken over university hospitals have so far tended not to interfere with their research programs. In some cases the chain has provided generous financial support. On their own some of the chains have taken advantage of their extensive hospital systems for applied research. HCA, for instance, has formed a subsidiary that coordinates HCA hospitals for clinical trials of new drugs, for which the company receives grants from pharmaceutical houses. Humana supports research on new medical technologies and procedures and trains physicians in their use at certain of its hospitals, which it has designated "Centers of Excellence."

Even if its practical applications are limited, highly visible research can confer prestige and status, for which profit-making corporations are as eager as any nonprofit teaching hospital seeking star teams of investigators. Prestige was certainly one motivation for Humana's support of William C. DeVries and his research on artificial hearts at the Humana Heart Institute International in Louisville, where DeVries found the financial backing that had eluded him during his previous association with the University of Utah. Such arrangements, beneficial as they may be to individual workers, do entail ethical complexities, many observers believe. They raise the possibility of conflicts of interest when clinical research is done at an institution that stands to profit directly from the outcome.

Perhaps the sharpest criticisms of the for-profit chains concern treatment of the medically indigent—those who lack medical insurance, are ineligible for government assistance and cannot pay hospital bills on their own. Most studies have confirmed one contention: that for-profit hospital chains have acquired and built hospitals in areas that have relatively few patients who are uninsured or are covered by Medicaid, which generally reimburses hospitals and doctors at lower rates than Medicare and private insurers.

The chains have also been accused of devoting a smaller percentage of their budgets to charity care and bad debts than nonprofit hospitals do and of eliminating services that are heavily drawn on by indigent patients, such as trauma and obstetrics. The evidence bearing on these charges is somewhat contradictory, and hypotheses and anecdotes abound. An analysis done in 1981 by the Office for Civil Rights of the U.S. Department of Health and Human Services considered hospitals in for-profit chains together with freestanding proprietary hospitals and found they admitted a slightly smaller proportion of uninsured patients than nonprofit hospitals. In 1983 the American Hospital Association took unpaid charges as a measure of charity care and found no statistically significant differences between for-profit and nonprofit hospitals.

The Institute of Medicine, however, examined data on uncompensated care in five states: California, Florida, Tennessee, Texas and Virginia. Except for California, those states have relatively few public hospitals, and their Medicaid eligibility requirements are restrictive. Hence the burden of un-



COSTS AND CHARGES in hospitals belonging to investor-owned chains (*color*) are compared with those in nonprofit hospitals. The comparison is based on data for pairs of hospitals, one hospital nonprofit and the other investor-owned, that were comparable in their location, size, type of service and other characteristics. The average daily cost of caring for a patient was higher in the investor-owned hospitals than it was in the nonprofit facilities, and the gap remained when the cost was divided into routine costs and ancillary costs—costs for tests and drugs, for example. The disparity suggests that the investor-owned hospitals operated less efficiently. Those hospitals also charged more than the nonprofit ones did, in particular for ancillary services, enabling the for-profit hospitals to make money in spite of their lower efficiency. The comparison, done by J. Michael Watt, Robert A. Derzon and James S. Hahn of Lewin and Associates, Inc., and a group from Johns Hopkins University, is based on figures from 1978 and 1980; since then the Government and private insurance companies have introduced cost-containment measures that place a premium on efficiency. A future study might well yield a different picture.

compensated care falls particularly heavily on voluntary and for-profit hospitals. Under those circumstances the two kinds of hospitals acquit themselves quite differently. In Florida, Texas and Virginia hospitals in private nonprofit chains spent almost twice as much on charity care as hospitals belonging to investor-owned chains. In Tennessee the nonprofit hospitals spent almost three times as much.

In none of the five states did private hospitals of either kind provide as much charity care as public hospitals, and the same disparity is seen in every other state. Nationwide, for-profit and nonprofit hospitals together devote only between 3 and 5 percent of their revenues to uncompensated care. In response to cost-cutting measures and the burden of surplus hospital beds both kinds of hospitals are attempting to reduce uncompensated care still further. They limit the number of uninsured patients who are admitted by setting quotas, by eliminating or cutting down on services that attract large numbers of the uninsured and by requiring payment in advance or doing "wallet biopsies." Such practices by nonprofit hospitals have led some observers to call for an end to their tax-exempt status.

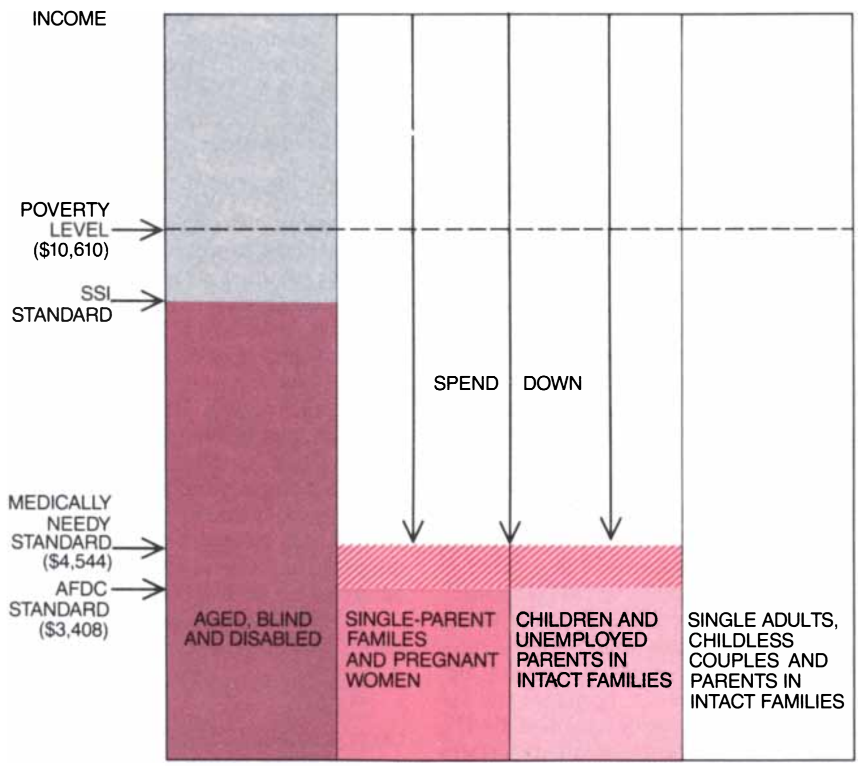
For-profit hospitals have sometimes compounded the problem by competing aggressively for paying patients, thereby depriving voluntary hospitals of the revenue needed to support uncompensated care. In Florida during the early 1980's, competition with for-profit hospitals put financial pressure on voluntary hospitals, which rejected increasing numbers of indigent patients. The burden of charity care fell more heavily than ever on the public hospitals, and because they operate on a fixed budget the burden soon threw them into financial crisis. The ensuing political crisis led to an expansion of the state's Medicaid program financed in part by an assessment on hospital revenues.

Such institutional failure exacts an appalling human cost. Private hospitals increasingly turn away indigent patients who arrive in their emergency rooms, often without first ensuring that a patient's condition is stable. A team of physicians at Cook County Hospital, a public hospital in Chicago, recently tracked 467 patients who were transferred there from the emergency rooms of for-profit and voluntary hospitals (including many nonprofit teaching hospitals). Follow-up data revealed that 24 percent of the patients were transferred in an unstable condition and 22 percent had to be admitted to an intensive-care unit on arrival. Of the nonsurgical patients who were transferred, 9.4 percent died during their hospitalization.

The investor-owned chains, through their own actions and the competition and commercialism they have spawned, draw attention to the fundamental problem: the U.S. stands as the only remaining industrialized nation, except for South Africa, that does not cover as a right of citizenship the medical expenses of anyone who becomes seriously ill. More than half of the poor are not protected by Medicaid because of restrictions on eligibility, and many more people who are not poor lack insurance because they have been laid off from their jobs or work in low-paying positions. All told, some 35 million Americans have no medical insurance for part or all of each year. Until a national solution is devised, a large part of the U.S. population will have nowhere to turn when faced with catastrophic medical bills.

By highlighting other shortcomings of the health-care system the for-profit chains are speeding its transformation. Historically the chains themselves will be seen as transitional institutions. They grew out of the generous reimbursement system sought by the medical profession, and they exploited the system deftly, teaching lessons in capital formation, management and marketing that are now expertly practiced by many of their nonprofit counterparts. In the process the chains hastened the crisis in health costs and the ensuing cost-containment measures.

Those measures in turn have led the chains to transform themselves into integrated systems offering comprehensive services and in many cases selling insurance to pay for them. In their new forms the for-profit systems will continue to grow, not only as an economic force but also as a political one. Increasingly they will have the power to shape the economic and regulatory framework of U.S. health care as well as to seek new opportunities within it.



MEDICARE AND MEDICAID cover the medical costs of fewer than half of the poor. The eligibility requirements shown here are those in effect in Florida; the income levels applied to a family of four in 1984. Medicare (*gray*) pays medical costs for the aged, blind and disabled regardless of income, and Medicaid (*color*) covers single-parent families and pregnant women whose income (after certain deductions) falls at or below the income standard for the Aid to Families with Dependent Children (AFDC) program. (Each state sets its own standard, which is usually well below the poverty level.) Medicare recipients are also eligible for Medicaid if their income drops below the standard for the Supplemental Security Income (SSI) program; Medicaid then reimburses costs for services not covered by Medicare. Florida, like many other states, extends Medicaid coverage to children and unemployed parents in intact families below the AFDC standard. It also has a "spend down" provision, under which certain families can claim Medicaid benefits if medical expenses reduce their income to the "medically needy" standard or below. Many of the poor in Florida and in other states must nonetheless depend on charity care at public and private hospitals. The chart is based on one prepared by Lewin and Associates, Inc.

Chaos

There is order in chaos: randomness has an underlying geometric form. Chaos imposes fundamental limits on prediction, but it also suggests causal relationships where none were previously suspected

by James P. Crutchfield, J. Doyne Farmer, Norman H. Packard and Robert S. Shaw

The great power of science lies in the ability to relate cause and effect. On the basis of the laws of gravitation, for example, eclipses can be predicted thousands of years in advance. There are other natural phenomena that are not as predictable. Although the movements of the atmosphere obey the laws of physics just as much as the movements of the planets do, weather forecasts are still stated in terms of probabilities. The weather, the flow of a mountain stream, the roll of the dice all have unpredictable aspects. Since there is no clear relation between cause and effect, such phenomena are said to have random elements. Yet until recently there was little reason to doubt that precise predictability could in principle be achieved. It was assumed that it was only necessary to gather and process a sufficient amount of information.

Such a viewpoint has been altered by a striking discovery: simple deterministic systems with only a few elements can generate random behavior. The randomness is fundamental; gathering more information does not make it go away. Randomness generated in this way has come to be called chaos.

A seeming paradox is that chaos is deterministic, generated by fixed rules that do not themselves involve any elements of chance. In principle the future is completely determined by the past, but in practice small uncertainties are amplified, so that even though the behavior is predictable in the short term, it is unpredictable in the long term. There is order in chaos: underlying chaotic behavior there are elegant geometric forms that create randomness in the same way as a card dealer shuffles a deck of cards or a blender mixes cake batter.

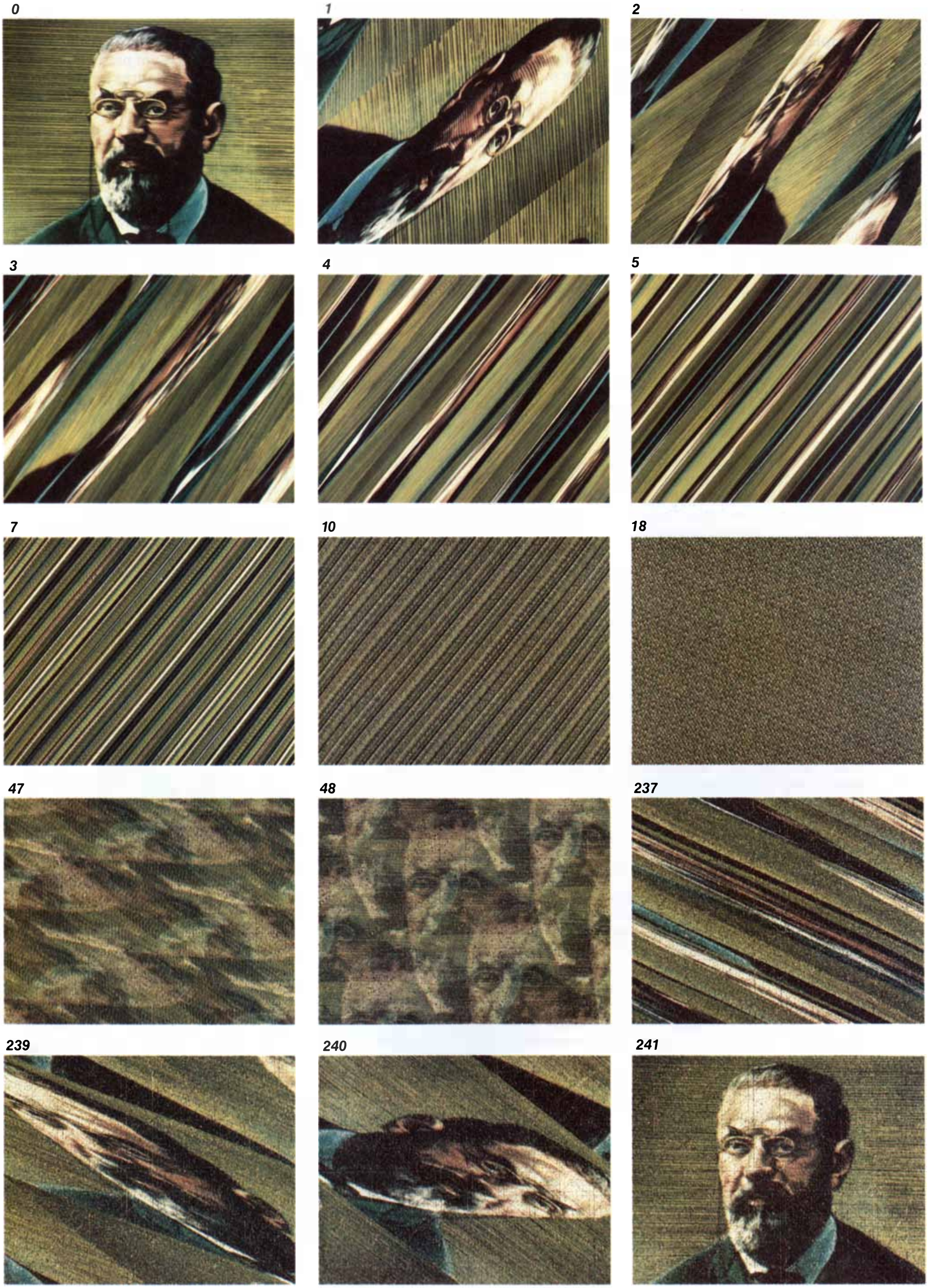
The discovery of chaos has created a new paradigm in scientific modeling. On one hand, it implies new fundamental limits on the ability to make predictions. On the other hand, the determinism inherent in chaos implies that many random phenomena are more predictable than had been thought. Random-looking information gathered in the past—and shelved because it was assumed to be too complicated—can now be explained in terms of simple laws. Chaos allows order to be found in such diverse systems as the atmosphere, dripping faucets and the heart. The result is a revolution that is affecting many different branches of science.

What are the origins of random behavior? Brownian motion provides a classic example of randomness. A speck of dust observed through a microscope is seen to move in a continuous and erratic jiggle. This is owing to the bombardment of the dust particle by the surrounding water molecules in thermal motion. Because the water molecules are unseen and exist in great number, the detailed motion of the dust particle is thoroughly unpredictable. Here the web of causal influences among the subunits can become so tangled that the resulting pattern of behavior becomes quite random.

The chaos to be discussed here requires no large number of subunits or unseen influences. The existence of random behavior in very simple systems motivates a reexamination of the sources of randomness even in large systems such as weather.

What makes the motion of the atmosphere so much harder to anticipate than the motion of the solar system? Both are made up of many parts, and both are governed by Newton's second law, $F = ma$, which can be viewed as a simple prescription for predicting the future. If the forces F acting on a given mass m are known, then so is the acceleration a. It then follows from the rules of calculus that if the position and velocity of an object can be measured at a given instant, they are determined forever. This is such a powerful idea that the 18th-century French mathematician Pierre Simon de Laplace once boasted that given the position and velocity of every particle in the universe, he could predict the future for the rest of time. Although there are several obvious practical difficulties to achieving Laplace's goal, for more than 100 years there seemed to be no reason for his not being right, at least in principle. The literal application of Laplace's dictum to human behavior led to the philosophical conclusion that human behavior

CHAOS results from the geometric operation of stretching. The effect is illustrated for a painting of the French mathematician Henri Poincaré, the originator of dynamical systems theory. The initial image (*top left*) was digitized so that a computer could perform the stretching operation. A simple mathematical transformation stretches the image diagonally as though it were painted on a sheet of rubber. Where the sheet leaves the box it is cut and reinserted on the other side, as is shown in panel *1*. (The number above each panel indicates how many times the transformation has been made.) Applying the transformation repeatedly has the effect of scrambling the face (*panels* 2–4). The net effect is a random combination of colors, producing a homogeneous field of green (*panels* 10 *and* 18). Sometimes it happens that some of the points come back near their initial locations, causing a brief appearance of the original image (*panels* 47–48, 239–241). The transformation shown here is special in that the phenomenon of "Poincaré recurrence" (as it is called in statistical mechanics) happens much more often than usual; in a typical chaotic transformation recurrence is exceedingly rare, occurring perhaps only once in the lifetime of the universe. In the presence of any amount of background fluctuations the time between recurrences is usually so long that all information about the original image is lost.


0
1
2
3
4
5
7
10
18
47
48
237
239
240
241

was completely predetermined: free will did not exist.

Twentieth-century science has seen the downfall of Laplacian determinism, for two very different reasons. The first reason is quantum mechanics. A central dogma of that theory is the Heisenberg uncertainty principle, which states that there is a fundamental limitation to the accuracy with which the position and velocity of a particle can be measured. Such uncertainty gives a good explanation for some random phenomena, such as radioactive decay. A nucleus is so small that the uncertainty principle puts a fundamental limit on the knowledge of its motion, and so it is impossible to gather enough information to predict when it will disintegrate.

The source of unpredictability on a large scale must be sought elsewhere, however. Some large-scale phenomena are predictable and others are not. The distinction has nothing to do with quantum mechanics. The trajectory of a baseball, for example, is inherently predictable; a fielder intuitively makes use of the fact every time he or she catches the ball. The trajectory of a flying balloon with the air rushing out of it, in contrast, is not predictable; the balloon lurches and turns erratically at times and places that are impossible to predict. The balloon obeys Newton's laws just as much as the baseball does; then why is its behavior so much harder to predict than that of the ball?

The classic example of such a dichotomy is fluid motion. Under some circumstances the motion of a fluid is laminar—even, steady and regular—and easily predicted from equations. Under other circumstances fluid motion is turbulent—uneven, unsteady and irregular—and difficult to predict. The transition from laminar to turbulent behavior is familiar to anyone who has been in an airplane in calm weather and then suddenly encountered a thunderstorm. What causes the essential difference between laminar and turbulent motion?

To understand fully why that is such a riddle, imagine sitting by a mountain stream. The water swirls and splashes as though it had a mind of its own, moving first one way and then another. Nevertheless, the rocks in the stream bed are firmly fixed in place, and the tributaries enter at a nearly constant rate of flow. Where, then, does the random motion of the water come from?

The late Soviet physicist Lev D. Landau is credited with an explanation of random fluid motion that held sway for many years, namely that the motion of a turbulent fluid contains many different, independent oscillations. As the fluid is made to move faster, causing it to become more turbulent, the oscillations enter the motion one at a time. Although each separate oscillation may be simple, the complicated combined motion renders the flow impossible to predict.

Landau's theory has been disproved, however. Random behavior occurs even in very simple systems, without any need for complication or indeterminacy. The French mathematician Henri Poincaré realized this at the turn of the century when he noted that unpredictable, "fortuitous" phenomena may occur in systems where a small change in the present causes a much larger change in the future. The notion is clear if one thinks of a rock poised at the top of a hill. A tiny push one way or another is enough to send it tumbling down widely differing paths. Although the rock is sensitive to small influences only at the top of the hill, chaotic systems are sensitive at every point in their motion.

A simple example serves to illustrate just how sensitive some physical

Laplace, 1776

"The present state of the system of nature is evidently a consequence of what it was in the preceding moment, and if we conceive of an intelligence which at a given instant comprehends all the relations of the entities of this universe, it could state the respective positions, motions, and general affects of all these entities at any time in the past or future.

"Physical astronomy, the branch of knowledge which does the greatest honor to the human mind, gives us an idea, albeit imperfect, of what such an intelligence would be. The simplicity of the law by which the celestial bodies move, and the relations of their masses and distances, permit analysis to follow their motions up to a certain point; and in order to determine the state of the system of these great bodies in past or future centuries, it suffices for the mathematician that their position and their velocity be given by observation for any moment in time. Man owes that advantage to the power of the instrument he employs, and to the small number of relations that it embraces in its calculations. But ignorance of the different causes involved in the production of events, as well as their complexity, taken together with the imperfection of analysis, prevents our reaching the same certainty about the vast majority of phenomena. Thus there are things that are uncertain for us, things more or less probable, and we seek to compensate for the impossibility of knowing them by determining their different degrees of likelihood. So it is that we owe to the weakness of the human mind one of the most delicate and ingenious of mathematical theories, the science of chance or probability."

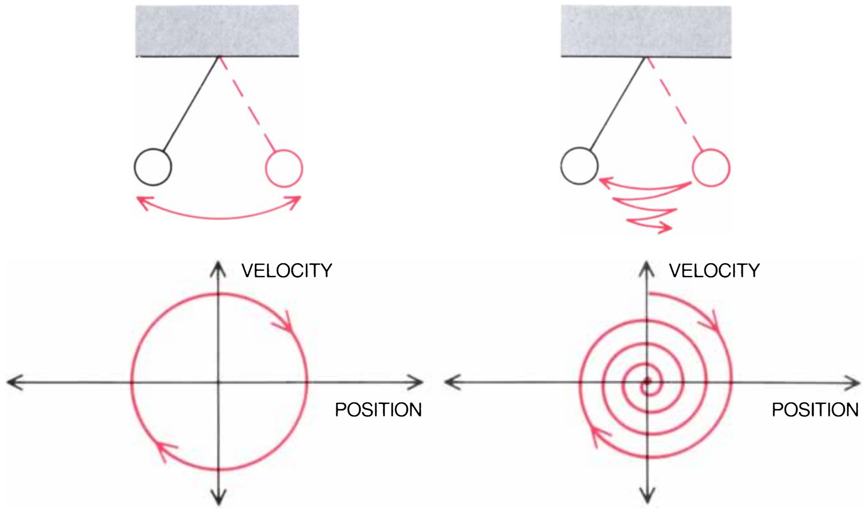
Poincaré, 1903

"A very small cause which escapes our notice determines a considerable effect that we cannot fail to see, and then we say that the effect is due to chance. If we knew exactly the laws of nature and the situation of the universe at the initial moment, we could predict exactly the situation of that same universe at a succeeding moment. But even if it were the case that the natural laws had no longer any secret for us, we could still only know the initial situation *approximately*. If that enabled us to predict the succeeding situation with *the same approximation*, that is all we require, and we should say that the phenomenon had been predicted, that it is governed by laws. But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon."

OUTLOOKS OF TWO LUMINARIES on chance and probability are contrasted. The French mathematician Pierre Simon de Laplace proposed that the laws of nature imply strict determinism and complete predictability, although imperfections in observations make the introduction of probabilistic theory necessary. The quotation from Poincaré foreshadows the contemporary view that arbitrarily small uncertainties in the state of a system may be amplified in time and so predictions of the distant future cannot be made.

systems can be to external influences. Imagine a game of billiards, somewhat idealized so that the balls move across the table and collide with a negligible loss of energy. With a single shot the billiard player sends the collection of balls into a protracted sequence of collisions. The player naturally wants to know the effects of the shot. For how long could a player with perfect control over his or her stroke predict the cue ball's trajectory? If the player ignored an effect even as minuscule as the gravitational attraction of an electron at the edge of the galaxy, the prediction would become wrong after one minute!

The large growth in uncertainty comes about because the balls are curved, and small differences at the point of impact are amplified with each collision. The amplification is exponential: it is compounded at every collision, like the successive reproduction of bacteria with unlimited space and food. Any effect, no matter how small, quickly reaches macroscopic proportions. That is one of the basic properties of chaos.



STATE SPACE is a useful concept for visualizing the behavior of a dynamical system. It is an abstract space whose coordinates are the degrees of freedom of the system's motion. The motion of a pendulum (*top*), for example, is completely determined by its initial position and velocity. Its state is thus a point in a plane whose coordinates are position and velocity (*bottom*). As the pendulum swings back and forth it follows an "orbit," or path, through the state space. For an ideal, frictionless pendulum the orbit is a closed curve (*bottom left*); otherwise, with friction, the orbit spirals to a point (*bottom right*).

It is the exponential amplification of errors due to chaotic dynamics that provides the second reason for Laplace's undoing. Quantum mechanics implies that initial measurements are always uncertain, and chaos ensures that the uncertainties will quickly overwhelm the ability to make predictions. Without chaos Laplace might have hoped that errors would remain bounded, or at least grow slowly enough to allow him to make predictions over a long period. With chaos, predictions are rapidly doomed to gross inaccuracy.

The larger framework that chaos emerges from is the so-called theory of dynamical systems. A dynamical system consists of two parts: the notions of a state (the essential information about a system) and a dynamic (a rule that describes how the state evolves with time). The evolution can be visualized in a state space, an abstract construct whose coordinates are the components of the state. In general the coordinates of the state space vary with the context; for a mechanical system they might be position and velocity, but for an ecological model they might be the populations of different species.

A good example of a dynamical system is found in the simple pendulum. All that is needed to determine its motion are two variables: position and velocity. The state is thus a point in a plane, whose coordinates are position and velocity. Newton's laws provide a rule, expressed mathematically as a differential equation, that describes how the state evolves. As the pendulum swings back and forth the state moves along an "orbit," or path, in the plane. In the ideal case of a frictionless pendulum the orbit is a loop; failing that, the orbit spirals to a point as the pendulum comes to rest.

A dynamical system's temporal evolution may happen in either continuous time or in discrete time. The former is called a flow, the latter a mapping. A pendulum moves continuously from one state to another, and so it is described by a continuous-time flow. The number of insects born each year in a specific area and the time interval between drops from a dripping faucet are more naturally described by a discrete-time mapping.

To find how a system evolves from a given initial state one can employ the dynamic (equations of motion) to move incrementally along an orbit. This method of deducing the system's behavior requires computational effort proportional to the desired length of time to follow the orbit. For simple systems such as a frictionless pendulum the equations of motion may occasionally have a closed-form solution, which is a formula that expresses any future state in terms of the initial state. A closed-form solution provides a short cut, a simpler algorithm that needs only the initial state and the final time to predict the future without stepping through intermediate states. With such a solution the algorithmic effort required to follow the motion of the system is roughly independent of the time desired. Given the equations of planetary and lunar motion and the earth's and moon's positions and velocities, for instance, eclipses may be predicted years in advance.

Success in finding closed-form solutions for a variety of simple systems during the early development of physics led to the hope that such solutions exist for any mechanical system. Unfortunately, it is now known that this is not true in general. The unpredictable behavior of chaotic dynamical systems cannot be expressed in a closed-form solution. Consequently there are no possible short cuts to predicting their behavior.

The state space nonetheless provides a powerful tool for describing the behavior of chaotic systems. The usefulness of the state-space picture lies in the ability to represent behavior in geometric form. For example, a pendulum that moves with friction eventually comes to a halt, which in the state space means the orbit approaches a point. The point does not move—it is a fixed point—and since it attracts nearby orbits, it is known as an attractor. If the pendulum is given a small push, it returns to the same fixed-point attractor. Any system that comes to rest with the passage of time can be characterized by a fixed point in state space. This is an example of a very general phenomenon, where losses due to friction or viscosity, for example, cause orbits to be attracted to a smaller region of the state space with lower dimension. Any such region is called an attractor. Roughly speaking, an attractor is what the behavior of a system settles down to, or is attracted to.

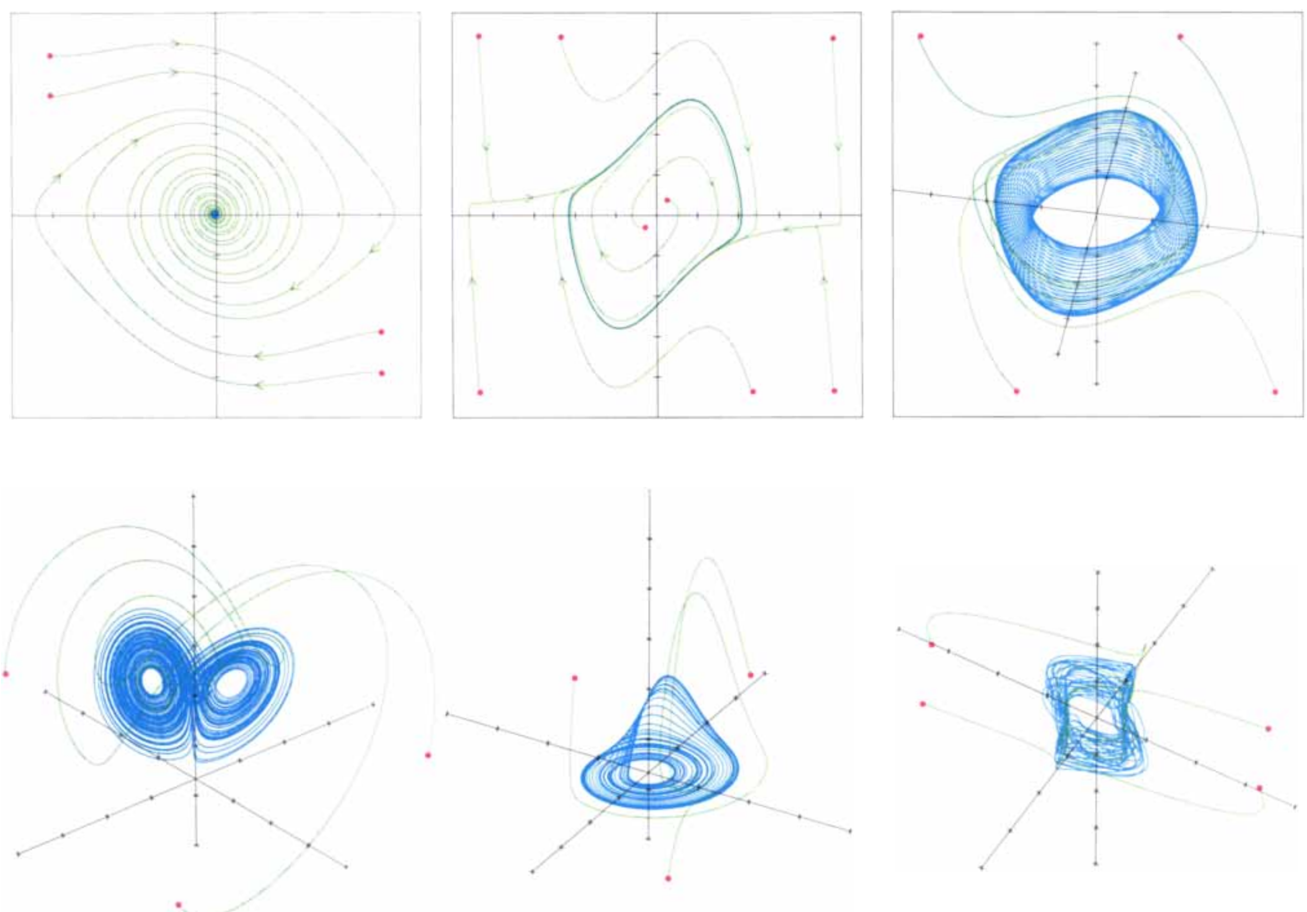
Some systems do not come to rest in the long term but instead cycle periodically through a sequence of states. An example is the pendulum clock, in which energy lost to friction is replaced by a mainspring or weights. The pendulum repeats the same motion over and over again. In the state space such a motion corresponds to a cycle, or periodic orbit. No matter how the pendulum is set swinging, the cycle approached in the long-term limit is the same. Such attractors are therefore called limit cycles. Another familiar system with a limit-cycle attractor is the heart.

A system may have several attractors. If that is the case, different initial conditions may evolve to different attractors. The set of points that evolve to an attractor is called its basin of attraction. The pendulum clock has two such basins: small displacements of the pendulum from its rest position result in a return to rest; with large displacements, however, the clock begins to tick as the pendulum executes a stable oscillation.

The next most complicated form of attractor is a torus, which resembles the surface of a doughnut. This shape describes motion made up of two independent oscillations, sometimes called quasi-periodic motion. (Physical examples can be constructed from driven electrical oscillators.) The orbit winds around the torus in state space, one frequency determined by how fast the orbit circles the doughnut in the short direction, the other regulated by how fast the orbit circles the long way around. Attractors may also be higher-dimensional tori, since they represent the combination of more than two oscillations.

The important feature of quasi-periodic motion is that in spite of its complexity it is predictable. Even though the orbit may never exactly repeat itself, if the frequencies that make up the motion have no common divisor, the motion remains regular. Orbits that start on the torus near one another remain near one another, and long-term predictability is guaranteed.

Until fairly recently, fixed points, limit cycles and tori were the only known attractors. In 1963 Edward N. Lorenz of the Massachusetts Institute



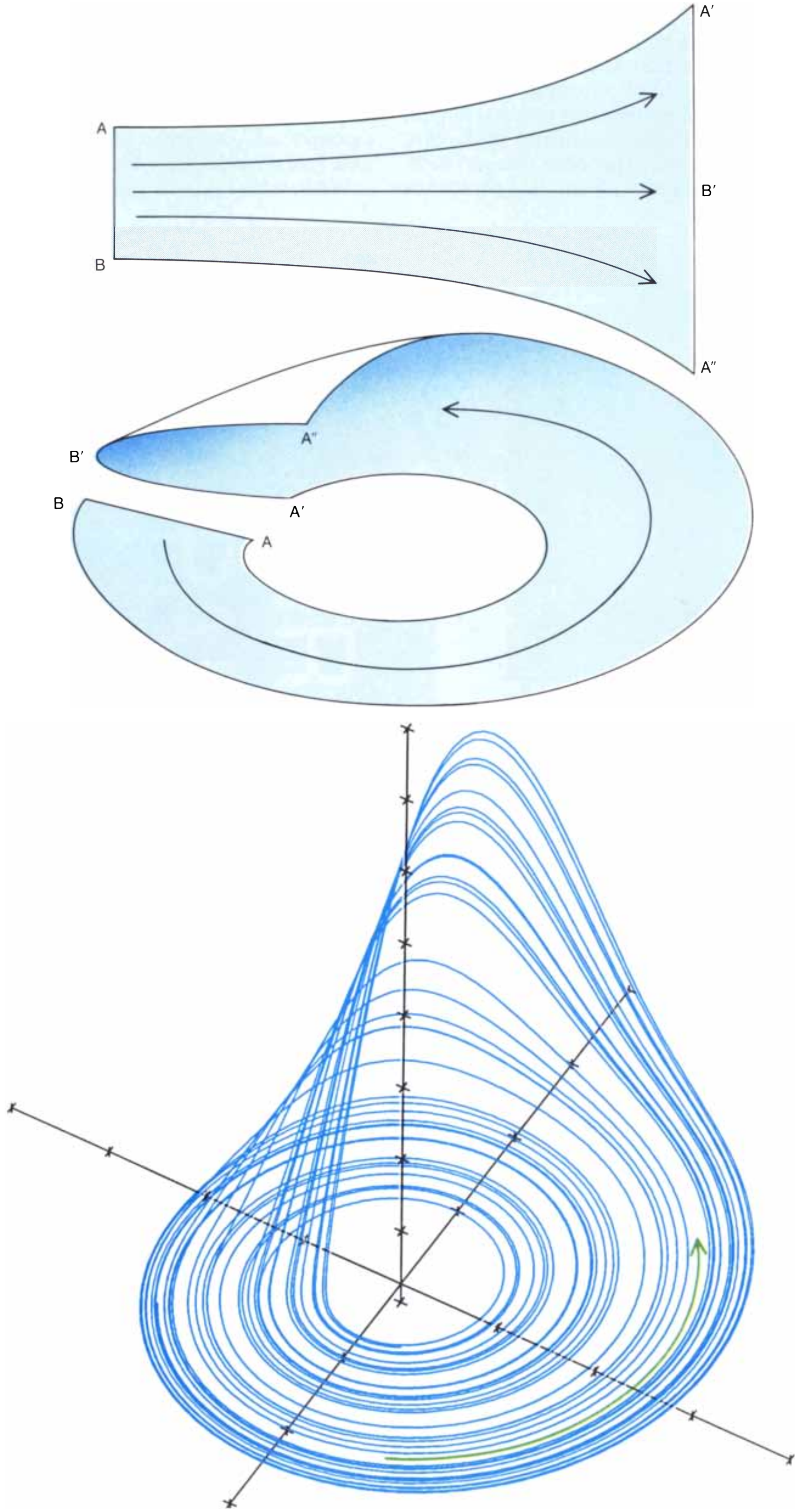
ATTRACTORS are geometric forms that characterize long-term behavior in the state space. Roughly speaking, an attractor is what the behavior of a system settles down to, or is attracted to. Here attractors are shown in blue and initial states in red. Trajectories (*green*) from the initial states eventually approach the attractors. The simplest kind of attractor is a fixed point (*top left*). Such an attractor corresponds to a pendulum subject to friction; the pendulum always comes to the same rest position, regardless of how it is started swinging (*see right half of illustration on preceding page*). The next most complicated attractor is a limit cycle (*top middle*), which forms a closed loop in the state space. A limit cycle describes stable oscillations, such as the motion of a pendulum clock and the beating of a heart. Compound oscillations, or quasi-periodic behavior, correspond to a torus attractor (*top right*). All three attractors are predictable: their behavior can be forecast as accurately as desired. Chaotic attractors, on the other hand, correspond to unpredictable motions and have a more complicated geometric form. Three examples of chaotic attractors are shown in the bottom row; from left to right they are the work of Edward N. Lorenz, Otto E. Rössler and one of the authors (Shaw) respectively. The images were prepared by using simple systems of differential equations having a three-dimensional state space.

of Technology discovered a concrete example of a low-dimensional system that displayed complex behavior. Motivated by the desire to understand the unpredictability of the weather, he began with the equations of motion for fluid flow (the atmosphere can be considered a fluid), and by simplifying them he obtained a system that had just three degrees of freedom. Nevertheless, the system behaved in an apparently random fashion that could not be adequately characterized by any of the three attractors then known. The attractor he observed, which is now known as the Lorenz attractor, was the first example of a chaotic, or strange, attractor.

Employing a digital computer to simulate his simple model, Lorenz elucidated the basic mechanism responsible for the randomness he observed: microscopic perturbations are amplified to affect macroscopic behavior. Two orbits with nearby initial conditions diverge exponentially fast and so stay close together for only a short time. The situation is qualitatively different for nonchaotic attractors. For these, nearby orbits stay close to one another, small errors remain bounded and the behavior is predictable.

The key to understanding chaotic behavior lies in understanding a simple stretching and folding operation, which takes place in the state space. Exponential divergence is a local feature: because attractors have finite size, two orbits on a chaotic attractor cannot diverge exponentially forever. Consequently the attractor must fold over onto itself. Although orbits diverge and follow increasingly different paths, they eventually must pass close to one another again. The orbits on a chaotic attractor are shuffled by this process, much as a deck of cards is shuffled by a dealer. The randomness of the chaotic orbits is the result of the shuffling process. The process of stretching and folding happens repeatedly, creating folds within folds ad infinitum. A chaotic attractor is, in other words, a fractal: an object that reveals more detail as it is increasingly magnified [*see illustration on page 53*].

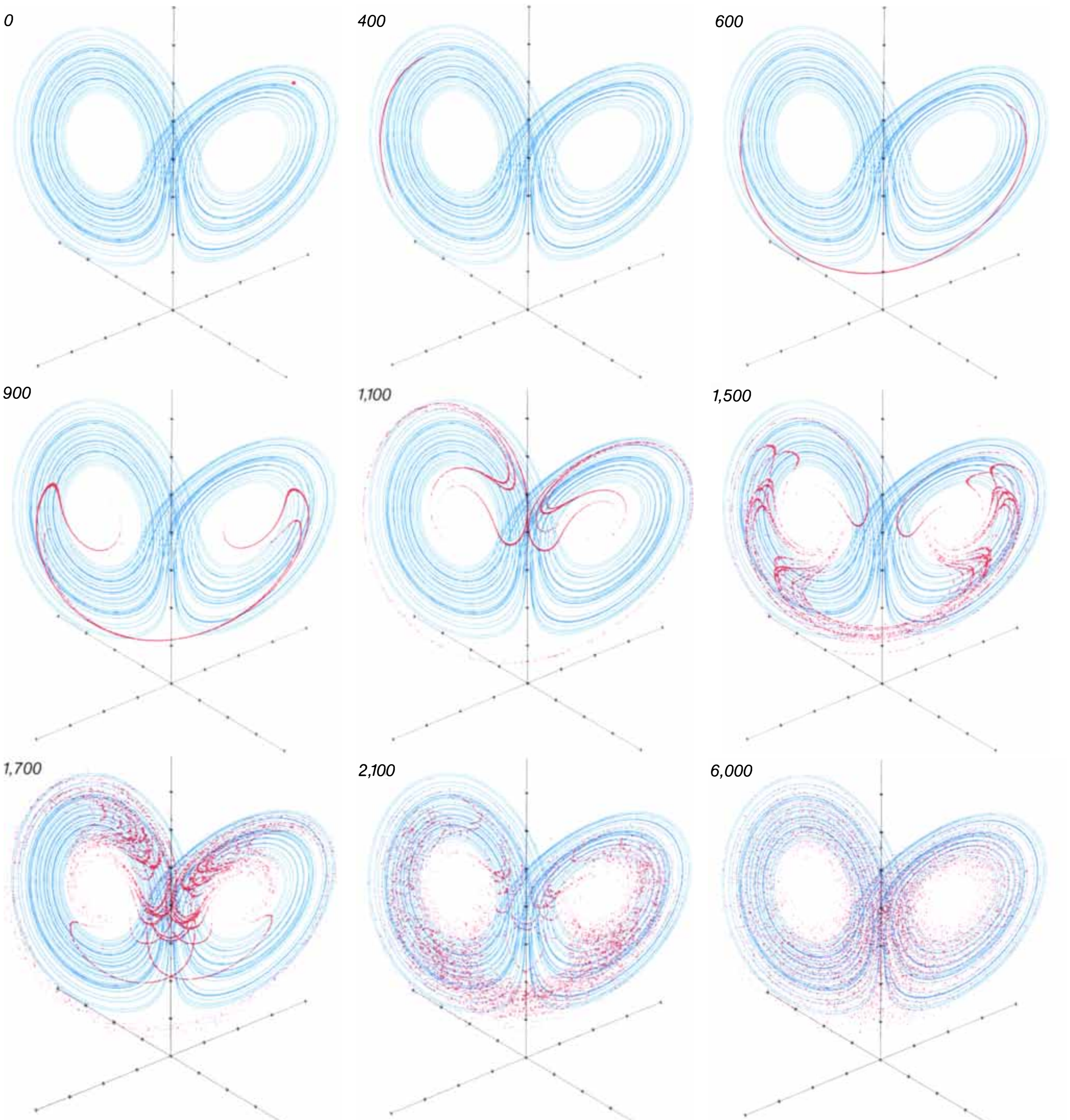
Chaos mixes the orbits in state space in precisely the same way as a baker mixes bread dough by kneading it. One can imagine what happens to nearby trajectories on a chaotic attractor by placing a drop of blue food coloring in the dough. The kneading is a combination of two actions: rolling out the dough, in which the food coloring is spread out, and folding the dough over. At first the blob of food coloring simply gets longer, but eventually it is folded, and after considerable time the blob is stretched and refolded many



CHAOTIC ATTRACTOR has a much more complicated structure than a predictable attractor such as a point, a limit cycle or a torus. Observed at large scales, a chaotic attractor is not a smooth surface but one with folds in it. The illustration shows the steps in making a chaotic attractor for the simplest case: the Rössler attractor (*bottom*). First, nearby trajectories on the object must "stretch," or diverge, exponentially (*top*); here the distance between neighboring trajectories roughly doubles. Second, to keep the object compact, it must "fold" back onto itself (*middle*): the surface bends onto itself so that the two ends meet. The Rössler attractor has been observed in many systems, from fluid flows to chemical reactions, illustrating Einstein's maxim that nature prefers simple forms.

times. On close inspection the dough consists of many layers of alternating blue and white. After only 20 steps the initial blob has been stretched to more than a million times its original length, and its thickness has shrunk to the molecular level. The blue dye is thoroughly mixed with the dough. Chaos works the same way, except that instead of mixing dough it mixes the state space. Inspired by this picture of mixing, Otto E. Rössler of the University of Tübingen created the simplest example of a chaotic attractor in a flow [*see illustration on preceding page*].

When observations are made on a physical system, it is impossible to specify the state of the system exactly owing to the inevitable errors in measurement. Instead the state of the system is located not at a single point but rather within a small region of state space. Although quantum uncertainty sets the ultimate size of the region, in



DIVERGENCE of nearby trajectories is the underlying reason chaos leads to unpredictability. A perfect measurement would correspond to a point in the state space, but any real measurement is inaccurate, generating a cloud of uncertainty. The true state might be anywhere inside the cloud. As shown here for the Lorenz attractor, the uncertainty of the initial measurement is represented by 10,000 red dots, initially so close together that they are indistinguishable. As each point moves under the action of the equations, the cloud is stretched into a long, thin thread, which then folds over onto itself many times, until the points are spread over the entire attractor. Prediction has now become impossible: the final state can be anywhere on the attractor. For a predictable attractor, in contrast, all the final states remain close together. The numbers above the illustrations are in units of 1/200 second.

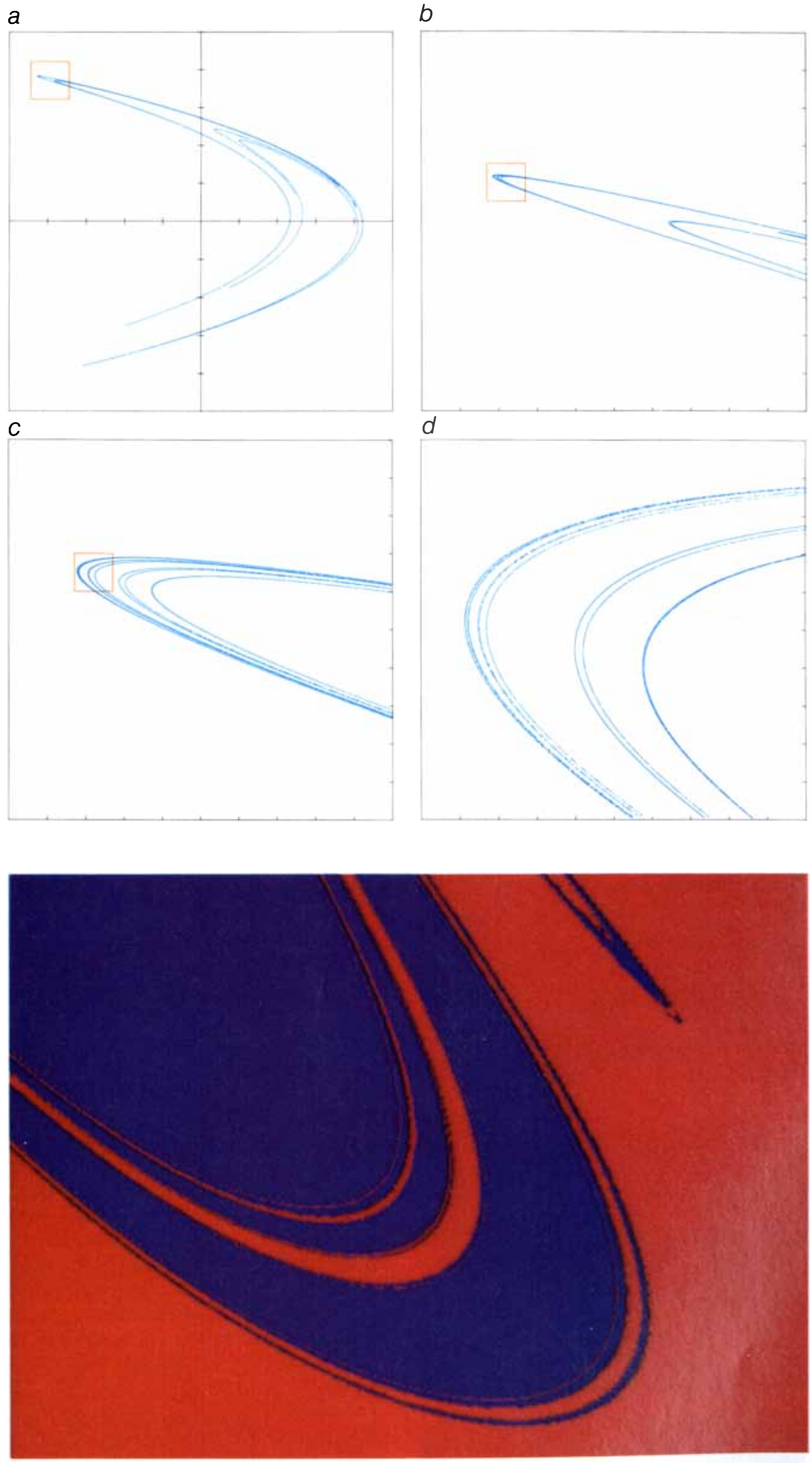
practice different kinds of noise limit measurement precision by introducing substantially larger errors. The small region specified by a measurement is analogous to the blob of blue dye in the dough.

Locating the system in a small region of state space by carrying out a measurement yields a certain amount of information about the system. The more accurate the measurement is, the more knowledge an observer gains about the system's state. Conversely, the larger the region, the more uncertain the observer. Since nearby points in nonchaotic systems stay close as they evolve in time, a measurement provides a certain amount of information that is preserved with time. This is exactly the sense in which such systems are predictable: initial measurements contain information that can be used to predict future behavior. In other words, predictable dynamical systems are not particularly sensitive to measurement errors.

The stretching and folding operation of a chaotic attractor systematically removes the initial information and replaces it with new information: the stretch makes small-scale uncertainties larger, the fold brings widely separated trajectories together and erases large-scale information. Thus chaotic attractors act as a kind of pump bringing microscopic fluctuations up to a macroscopic expression. In this light it is clear that no exact solution, no short cut to tell the future, can exist. After a brief time interval the uncertainty specified by the initial measurement covers the entire attractor and all predictive power is lost: there is simply no causal connection between past and future.

Chaotic attractors function locally as noise amplifiers. A small fluctuation due perhaps to thermal noise will cause a large deflection in the orbit position soon afterward. But there is an important sense in which chaotic attractors differ from simple noise amplifiers. Because the stretching and folding operation is assumed to be repetitive and continuous, any tiny fluctuation will eventually dominate the motion, and the qualitative behavior is independent of noise level. Hence chaotic systems cannot directly be "quieted," by lowering the temperature, for example. Chaotic systems generate randomness on their own without the need for any external random inputs. Random behavior comes from more than just the amplification of errors and the loss of the ability to predict; it is due to the complex orbits generated by stretching and folding.

It should be noted that chaotic as



CHAOTIC ATTRACTORS are fractals: objects that reveal more detail as they are increasingly magnified. Chaos naturally produces fractals. As nearby trajectories expand they must eventually fold over close to one another for the motion to remain finite. This is repeated again and again, generating folds within folds, ad infinitum. As a result chaotic attractors have a beautiful microscopic structure. Michel Hénon of the Nice Observatory in France discovered a simple rule that stretches and folds the plane, moving each point to a new location. Starting from a single initial point, each successive point obtained by repeatedly applying Hénon's rule is plotted. The resulting geometric form (*a*) provides a simple example of a chaotic attractor. The small box is magnified by a factor of 10 in *b*. By repeating the process (*c, d*) the microscopic structure of the attractor is revealed in detail. The bottom illustration depicts another part of the Hénon attractor.

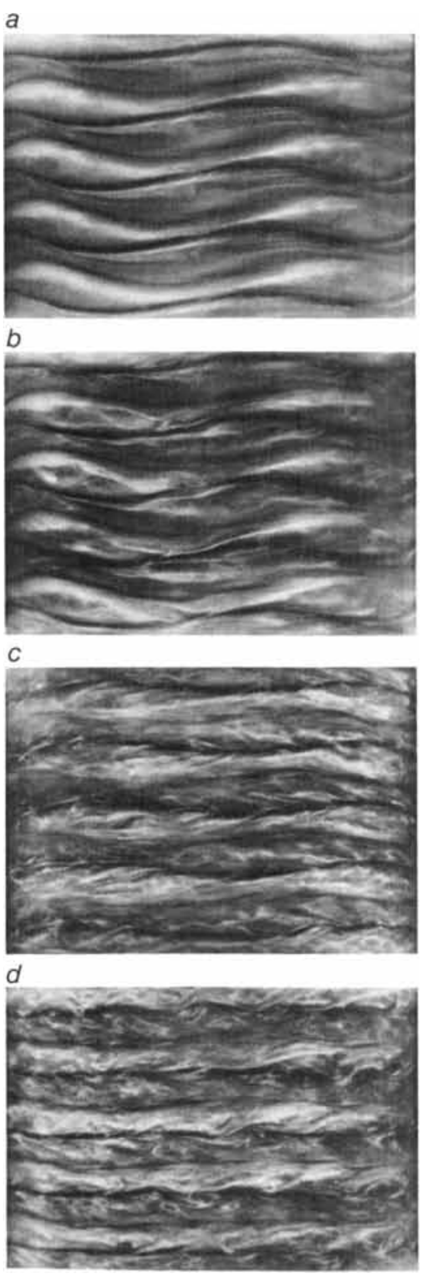
well as nonchaotic behavior can occur in dissipationless, energy-conserving systems. Here orbits do not relax onto an attractor but instead are confined to an energy surface. Dissipation is, however, important in many if not most real-world systems, and one can expect the concept of attractor to be generally useful.

Low-dimensional chaotic attractors open a new realm of dynamical systems theory, but the question remains of whether they are relevant to randomness observed in physical systems. The first experimental evidence supporting the hypothesis that chaotic attractors underlie random motion in fluid flow was rather indirect. The experiment was done in 1974 by Jerry P. Gollub of Haverford College and Harry L. Swinney of the University of Texas at Austin. The evidence was indirect because the investigators focused not on the attractor itself but rather on statistical properties characterizing the attractor.

The system they examined was a Couette cell, which consists of two concentric cylinders. The space between the cylinders is filled with a fluid, and one or both cylinders are rotated with a fixed angular velocity. As the angular velocity increases, the fluid shows progressively more complex flow patterns, with a complicated time dependence [*see illustration on this page*]. Gollub and Swinney essentially measured the velocity of the fluid at a given spot. As they increased the rotation rate, they observed transitions from a velocity that is constant in time to a periodically varying velocity and finally to an aperiodically varying velocity. The transition to aperiodic motion was the focus of the experiment.

The experiment was designed to distinguish between two theoretical pictures that predicted different scenarios for the behavior of the fluid as the rotation rate of the fluid was varied. The Landau picture of random fluid motion predicted that an ever higher number of independent fluid oscillations should be excited as the rotation rate is increased. The associated attractor would be a high-dimensional torus. The Landau picture had been challenged by David Ruelle of the Institut des Hautes Études Scientifiques near Paris and Floris Takens of the University of Groningen in the Netherlands. They gave mathematical arguments suggesting that the attractor associated with the Landau picture would not be likely to occur in fluid motion. Instead their results suggested that any possible high-dimensional tori might give way to a chaotic attractor, as originally postulated by Lorenz.

Gollub and Swinney found that for low rates of rotation the flow of the fluid did not change in time: the underlying attractor was a fixed point. As the rotation was increased the water began to oscillate with one independent frequency, corresponding to a limit-cycle attractor (a periodic orbit), and as the rotation was increased still further the oscillation took on two independent frequencies, corresponding to a two-dimensional torus attractor. Landau's theory predicted that as the rotation rate was further increased the pattern would continue: more distinct frequencies would gradually appear. Instead, at a critical rotation rate a continuous range of frequencies suddenly appeared. Such an observation was consistent with Lorenz' "deterministic nonperiodic flow," lending credence to his idea that chaotic attractors underlie fluid turbulence.



EXPERIMENTAL EVIDENCE supports the hypothesis that chaotic attractors underlie some kinds of random motion in fluid flow. Shown here are successive pictures of water in a Couette cell, which consists of two nested cylinders. The space between the cylinders is filled with water and the inner cylinder is rotated with a certain angular velocity (*a*). As the angular velocity is increased, the fluid shows a progressively more complex flow pattern (*b*), which becomes irregular (*c*) and then chaotic (*d*).

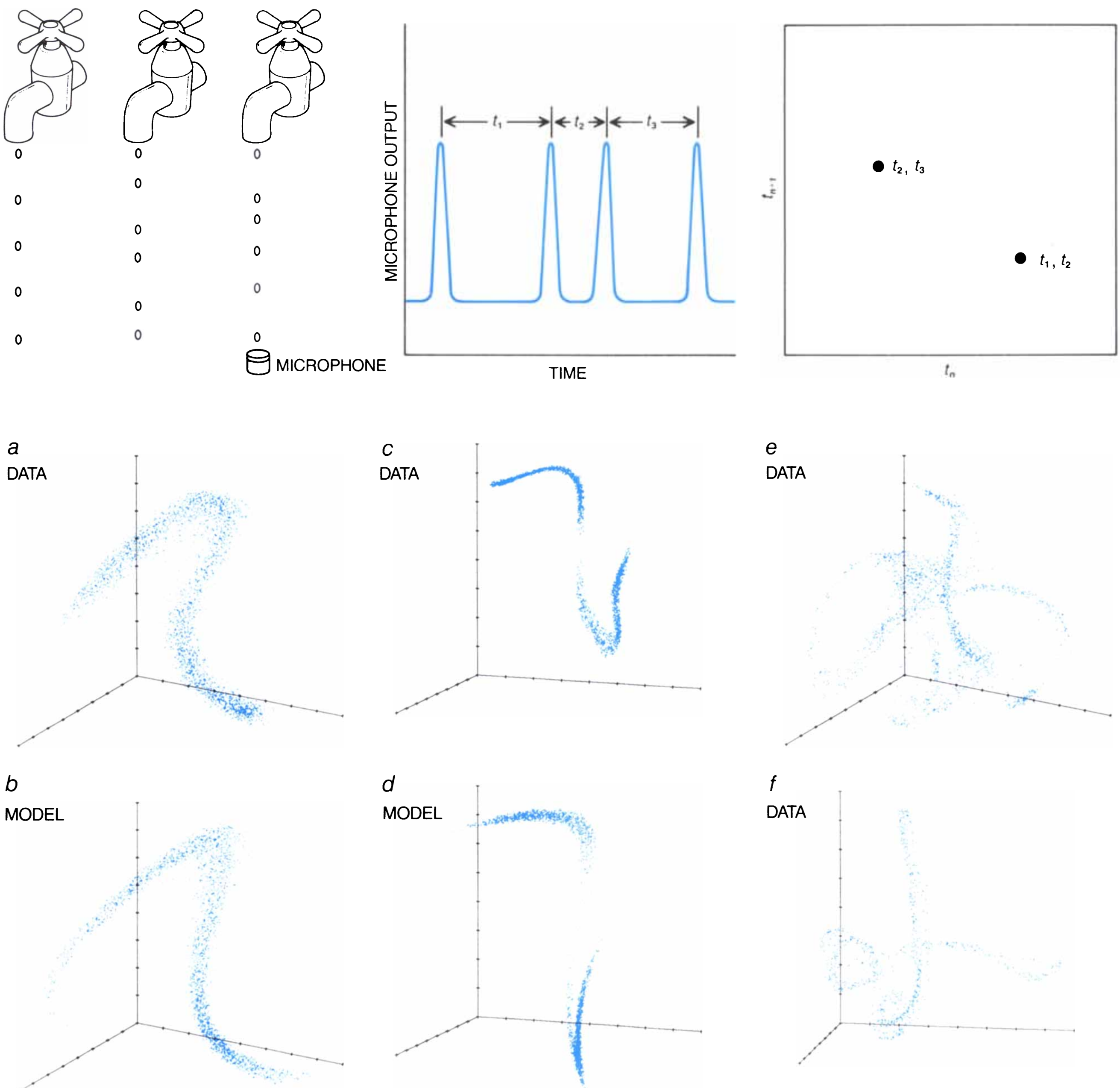
Although the analysis of Gollub and Swinney bolstered the notion that chaotic attractors might underlie some random motion in fluid flow, their work was by no means conclusive. One would like to explicitly demonstrate the existence in experimental data of a simple chaotic attractor. Typically, however, an experiment does not record all facets of a system but only a few. Gollub and Swinney could not record, for example, the entire Couette flow but only the fluid velocity at a single point. The task of the investigator is to "reconstruct" the attractor from the limited data. Clearly that cannot always be done; if the attractor is too complicated, something will be lost. In some cases, however, it is possible to reconstruct the dynamics on the basis of limited data.

A technique introduced by us and put on a firm mathematical foundation by Takens made it possible to reconstruct a state space and look for chaotic attractors. The basic idea is that the evolution of any single component of a system is determined by the other components with which it interacts. Information about the relevant components is thus implicitly contained in the history of any single component. To reconstruct an "equivalent" state space, one simply looks at a single component and treats the measured values at fixed time delays (one second ago, two seconds ago and so on, for example) as though they were new dimensions.

The delayed values can be viewed as new coordinates, defining a single point in a multidimensional state space. Repeating the procedure and taking delays relative to different times generates many such points. One can then use other techniques to test whether or not these points lie on a chaotic attractor. Although this representation is in many respects arbitrary, it turns out that the important properties of an attractor are preserved by it and do not depend on the details of how the reconstruction is done.

The example we shall use to illustrate the technique has the advantage of being familiar and accessible to nearly everyone. Most people are aware of the periodic pattern of drops emerging from a dripping faucet. The time between successive drops can be quite regular, and more than one insomniac has been kept awake waiting for the next drop to fall. Less familiar is the behavior of a faucet at a somewhat higher flow rate. One can often find a regime where the drops, while still falling separately, fall in a never repeating patter, like an infinitely inventive drummer. (This is an experiment easily carried out personally; the faucets without the little screens work best.) The changes between periodic and random-seeming patterns are reminiscent of the transition between laminar and turbulent fluid flow. Could a simple chaotic attractor underlie this randomness?

The experimental study of a dripping faucet was done at the University of California at Santa Cruz by one of



DRIPPING FAUCET is an example of a common system that can undergo a chaotic transition. The underlying attractor is reconstructed by plotting the time intervals between successive drops in pairs, as is shown at the top of the illustration. Attractors reconstructed from an actual dripping faucet (*a, c*) compare favorably with attractors generated by following variants of Hénon's rule (*b, d*). (The entire Hénon attractor is shown on page 53.) Illustrations *e* and *f* were reconstructed from high rates of water flow and presumably represent the cross sections of hitherto unseen chaotic attractors. Time-delay coordinates were employed in each of the plots. The horizontal coordinate is t_n, the time interval between drop *n* and drop *n* − 1. The vertical coordinate is the next time interval, t_{n+1}, and the third coordinate, visualized as coming out of the page, is t_{n+2}. Each point is thus determined by a triplex of numbers (t_n, t_{n+1}, t_{n+2}) that have been plotted for a set of 4,094 data samples. Simulated noise was added to illustrations *b* and *d*.

us (Shaw) in collaboration with Peter L. Scott, Stephen C. Pope and Philip J. Martein. The first form of the experiment consisted in allowing the drops from an ordinary faucet to fall on a microphone and measuring the time intervals between the resulting sound pulses. Typical results from a somewhat more refined experiment are shown on the preceding page. By plotting the time intervals between drops in pairs, one effectively takes a cross section of the underlying attractor. In the periodic regime, for example, the meniscus where the drops are detaching is moving in a smooth, repetitive manner, which could be represented by a limit cycle in the state space. But this smooth motion is inaccessible in the actual experiment; all that is recorded is the time intervals between the breaking off of the individual drops. This is like applying a stroboscopic light to regular motion around a loop. If the timing is right, one sees only a fixed point.

The exciting result of the experiment was that chaotic attractors were indeed found in the nonperiodic regime of the dripping faucet. It could have been the case that the randomness of the drops was due to unseen influences, such as small vibrations or air currents. If that was so, there would be no particular relation between one interval and the next, and the plot of the data taken in pairs would have shown only a featureless blob. The fact that any structure at all appears in the plots shows the randomness has a deterministic underpinning. In particular, many data sets show the horseshoelike shape that is the signature of the simple stretching and folding process discussed above. The characteristic shape can be thought of as a "snapshot" of a fold in progress, for example, a cross section partway around the Rössler attractor shown on page 51. Other data sets seem more complicated; these may be cross sections of higher-dimensional attractors. The geometry of attractors above three dimensions is almost completely unknown at this time.

If a system is chaotic, how chaotic is it? A measure of chaos is the "entropy" of the motion, which roughly speaking is the average rate of stretching and folding, or the average rate at which information is produced. Another statistic is the "dimension" of the attractor. If a system is simple, its behavior should be described by a low-dimensional attractor in the state space, such as the examples given in this article. Several numbers may be required to specify the state of a more complicated system, and its corresponding attractor would therefore be higher-dimensional.

The technique of reconstruction, combined with measurements of entropy and dimension, makes it possible to reexamine the fluid flow originally studied by Gollub and Swinney. This was done by members of Swinney's group in collaboration with two of us (Crutchfield and Farmer). The reconstruction technique enabled us to make images of the underlying attractor. The images do not give the striking demonstration of a low-dimensional attractor that studies of other systems, such as the dripping faucet, do. Measurements of the entropy and dimension reveal, however, that irregular fluid motion near the transition in Couette flow can be described by chaotic attractors. As the rotation rate of the Couette cell increases so do the entropy and dimension of the underlying attractors.

In the past few years a growing number of systems have been shown to exhibit randomness due to a simple chaotic attractor. Among them are the convection pattern of fluid heated in a small box, oscillating concentration levels in a stirred-chemical reaction, the beating of chicken-heart cells and a large number of electrical and mechanical oscillators. In addition computer models of phenomena ranging from epidemics to the electrical activity of a nerve cell to stellar oscillations have been shown to possess this simple type of randomness. There are even experiments now under way that are searching for chaos in areas as disparate as brain waves and economics.

It should be emphasized, however, that chaos theory is far from a panacea. Many degrees of freedom can also make for complicated motions that are effectively random. Even though a given system may be known to be chaotic, the fact alone does not reveal very much. A good example is molecules bouncing off one another in a gas. Although such a system is known to be chaotic, that in itself does not make prediction of its behavior easier. So many particles are involved that all that can be hoped for is a statistical description, and the essential statistical properties can be derived without taking chaos into account.

There are other uncharted questions for which the role of chaos is unknown. What of constantly changing patterns that are spatially extended, such as the dunes of the Sahara and fully developed turbulence? It is not clear whether complex spatial patterns can be usefully described by a single attractor in a single state space. Perhaps, though, experience with the simplest attractors can serve as a guide to a more advanced picture, which may involve entire assemblages of spatially mobile deterministic forms akin to chaotic attractors.

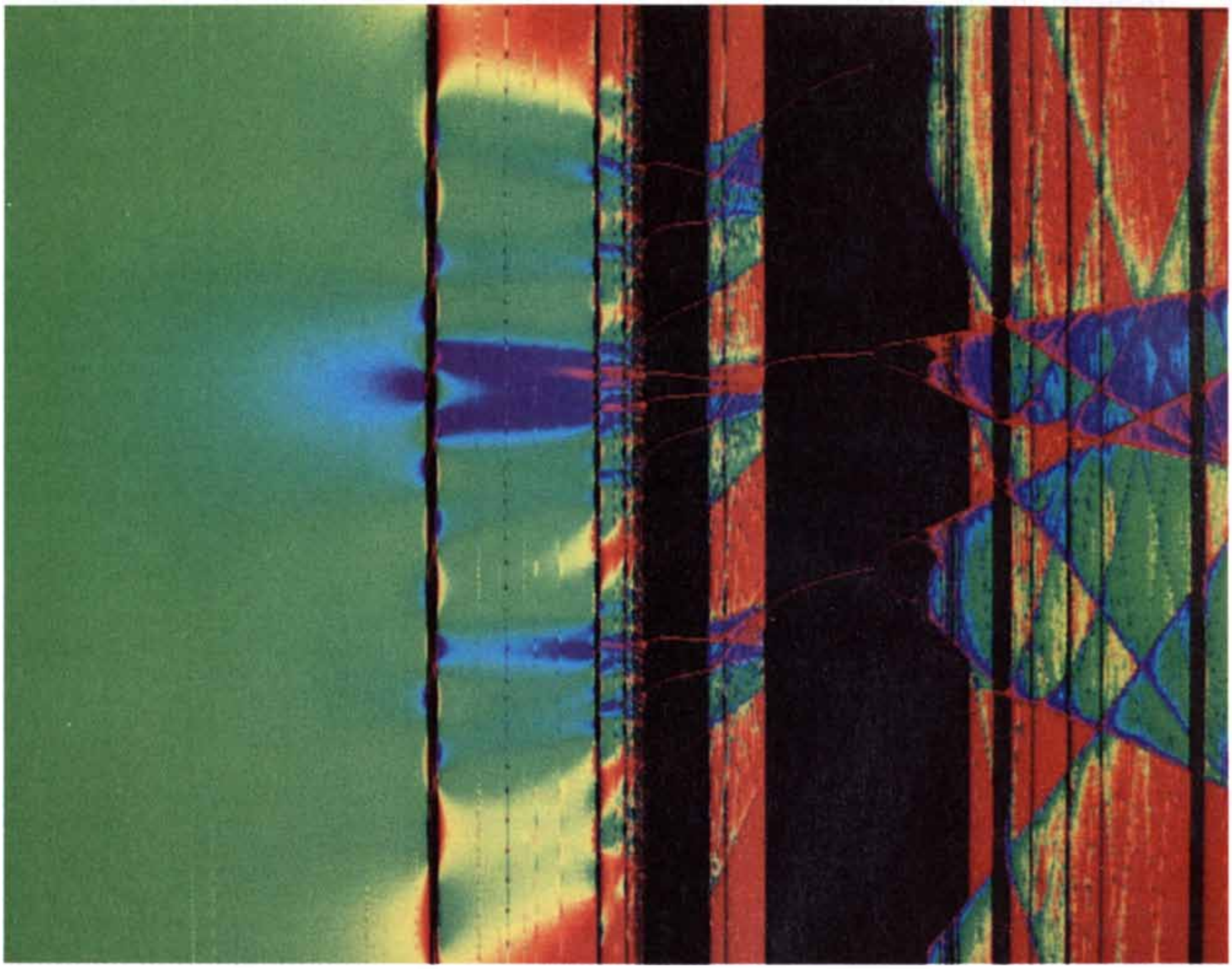
The existence of chaos affects the scientific method itself. The classic approach to verifying a theory is to make predictions and test them against experimental data. If the phenomena are chaotic, however, long-term predictions are intrinsically impossible. This has to be taken into account in judging the merits of the theory. The process of verifying a theory thus becomes a much more delicate operation, relying on statistical and geometric properties rather than on detailed prediction.

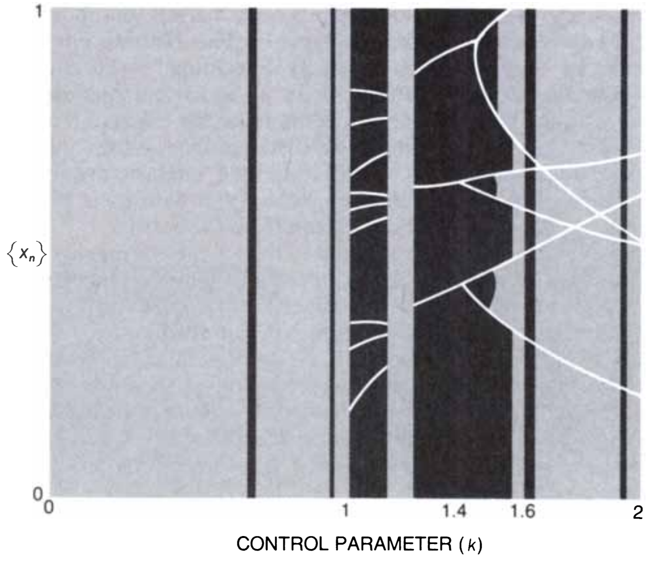
Chaos brings a new challenge to the reductionist view that a system can be understood by breaking it down and studying each piece. This view has been prevalent in science in part because there are so many systems for which the behavior of the whole is indeed the sum of its parts. Chaos demonstrates, however, that a system can have complicated behavior that emerges as a consequence of simple, nonlinear interaction of only a few components.

The problem is becoming acute in a wide range of scientific disciplines, from describing microscopic physics to modeling macroscopic behavior of biological organisms. The ability to obtain detailed knowledge of a system's structure has undergone a tremendous advance in recent years, but the ability to integrate this knowledge has been stymied by the lack of a proper conceptual framework within which to describe qualitative behavior. For example, even with a complete map of the nervous system of a simple organism, such as the nematode studied by Sidney Brenner of the University of Cambridge, the organism's behavior cannot be deduced. Similarly, the hope that physics could be complete with an increasingly detailed understanding of fundamental physical forces and constituents is unfounded. The interaction of components on one scale can lead to complex global behavior on a larger scale that in general cannot be deduced from knowledge of the individual components.

Chaos is often seen in terms of the limitations it implies, such as lack of predictability. Nature may, however, employ chaos constructively. Through amplification of small fluctuations it can provide natural systems with access to novelty. A prey escaping a predator's attack could use chaotic flight control as an element of surprise to evade capture. Biological evolution demands genetic variability; chaos provides a means of structuring random changes, thereby providing the possibility of putting variability under evolutionary control.

Even the process of intellectual progress relies on the injection of new ideas and on new ways of connecting old ideas. Innate creativity may have an underlying chaotic process that selectively amplifies small fluctuations and molds them into macroscopic coherent mental states that are experienced as thoughts. In some cases the thoughts may be decisions, or what are perceived to be the exercise of will. In this light, chaos provides a mechanism that allows for free will within a world governed by deterministic laws.





TRANSITION TO CHAOS is depicted schematically by means of a bifurcation diagram: a plot of a family of attractors (*vertical axis*) versus a control parameter (*horizontal axis*). The diagram was generated by a simple dynamical system that maps one number to another. The dynamical system used here is called a circle map, which is specified by the iterative equation $x_{n+1} = \omega + x_n + k/2\pi \cdot \sin(2\pi x_n)$. For each chosen value of the control parameter k a computer plotted the corresponding attractor. The colors encode the probability of finding points on the attractors: red corresponds to regions that are visited frequently, green to regions that are visited less frequently and blue to regions that are rarely visited. As k is increased from 0 to 2 (*see drawing at left*), the diagram shows two paths to chaos: a quasi-periodic route (from $k = 0$ to $k = 1$, which corresponds to the green region above) and a "period doubling" route (from $k = 1.4$ to $k = 2$). The quasi-periodic route is mathematically equivalent to a path that passes through a torus attractor. In the period-doubling route, which is based on the limit-cycle attractor, branches appear in pairs, following the geometric series 2, 4, 8, 16, 32 and so on. The iterates oscillate among the pairs of branches. (At a particular value of k—1.6, for instance—the iterates visit only two values.) Ultimately the branch structure becomes so fine that a continuous band structure emerges: a threshold is reached beyond which chaos appears.

Dark Matter in the Universe

More matter exists than is seen. The motions of stars and galaxies indicate where some of it is; theory suggests there is far more. What and where is it? Particle physics and astrophysics are yielding clues

by Lawrence M. Krauss

What is the universe made of? What kind of matter is commonest, how much is there and how is it distributed? These questions, always a focus of cosmology, have become even more intriguing over the past few years as evidence has piled up to support the proposition that most of the mass in the universe is dark—invisible to any existing telescope or other observational device—and new developments in both high-energy physics and astrophysics have made possible new predictions of the makeup and distribution of this possibly exotic form of matter.

There is already overwhelming evidence that the visible matter within galaxies may account for less than 10 percent of the galaxies' actual mass: the rest, not yet directly detectable by observers on the earth, is probably distributed within and around each galaxy. Theoretical considerations now suggest this may be only the tip of the cosmic "iceberg" of dark matter: much greater amounts of dark matter may be distributed throughout the universe, perhaps in configurations entirely independent of the distribution of galaxies. It may be that this mass can be accounted for only by the existence of new kinds of matter.

The question of dark matter—how much of it there is, how it is distributed and what it is made of—is intimately linked to questions about the overall structure and evolution of the universe: because dark matter is probably the dominant form of mass in the universe, it must have affected the evolution of the features observable today. Questions of structure in turn depend for their answers on a deep bond that has formed between macrophysics and microphysics, the bodies of knowledge that respectively describe interactions on the largest scale (that of the universe as a whole) and the smallest scale (that of the fundamental particles that make up all matter).

This bond is provided by the observation that the universe is expanding. If we are bold enough to extrapolate the expansion backward by between 10 and 20 billion years, the cosmological and microscopic scales begin to merge, because at the earliest times those structures now observed on the largest scales occupied regions having characteristic distances and energies on scales that are typically associated with the processes governing the interactions of fundamental particles. Since the structure remaining on the largest scales observable today reflects the imprint of those processes, it is natural to expect the resolution of the dark-matter question to come in part from advances in the understanding of the physics of high-energy particles.

At present a number of testable predictions for the nature of both the dark matter and the primordial structures in the early universe have been proposed. Future developments, both theoretical and observational, will help to decide issues ranging from how and when galaxies and stars first formed to what kinds of symmetries underlie the interactions of particles at very high energies. Ultimately the debate about dark matter may help to answer a question as old as human inquiry: What will be the fate of the universe?

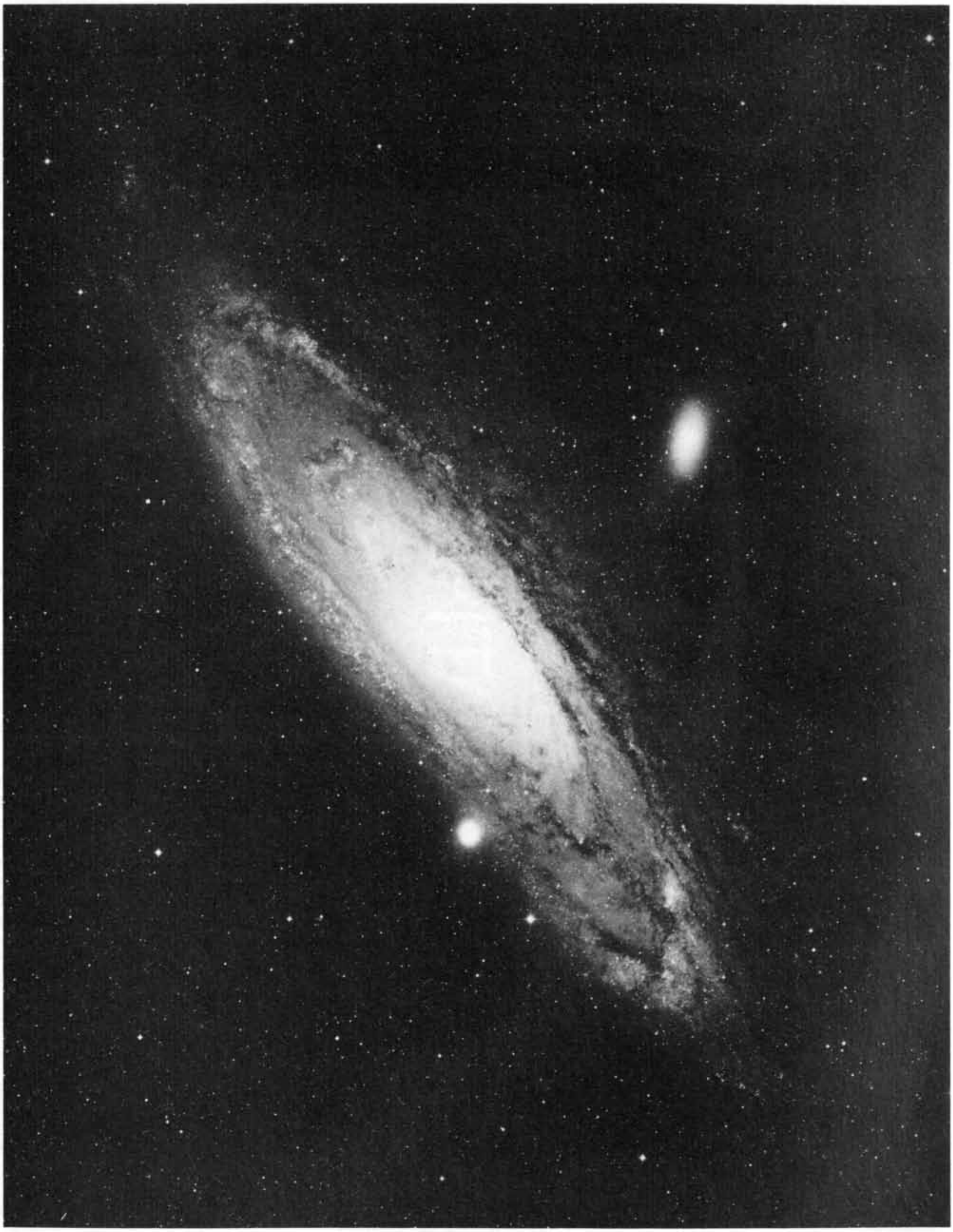
Ever since the early 1930's, when Edwin P. Hubble confirmed that the universe is expanding, it has been natural to ask whether the expansion will eventually halt. The answer depends on two factors: how fast the universe is currently expanding and how strongly the force of gravity, determined by the average density of mass within the universe, holds that mass together. A high mass density would cause a strong gravitational attraction.

According to the general theory of relativity, there is a relation between the magnitudes of these two factors and the mean curvature of the universe [*see illustration on page 60*]. If the average mass density is high enough to halt the expansion and cause the universe to contract again, the universe is said to be closed. If the density is so small compared with the expansion rate that the universe will continue to expand at a finite rate forever, the universe is said to be open. If the gravitational attraction is precisely strong enough to continue to slow the expansion but not strong enough to close the universe, the universe is said to be flat.

Because the observable universe is highly uniform in all directions, its rate of expansion can be described in terms of a single parameter, which is known as the Hubble constant even though it is actually a slowly varying function of time. The Hubble constant is the average speed with which any two regions of the universe are moving apart from each other divided by the distance between them.

For any given measurement of the Hubble constant, it is easy to determine the mass density that would correspond to a flat universe. Measurements of the Hubble constant, however, depend on a variety of uncertain measurements. The Hubble constant is generally determined by measuring the velocity at which various objects are receding from the earth and gauging their distance by such techniques as estimating their intrinsic brightness and comparing that with their brightness as seen from the earth.

Because those measurements are highly uncertain, there is a spread of about a factor of two in current determinations of the universe's rate of expansion. As an upper limit, objects one megaparsec (about 3.26 million light-years) apart are on the average receding from one another at a speed somewhat less than about 100 kilometers per second. At that rate the average mass density that would result in a flat universe is about 2×10^{-29} gram per cubic centimeter, which is roughly



SPIRAL GALAXY M31 (ANDROMEDA) reveals the presence of dark matter by the motion of its outer arms. They are rotating about the galactic center faster than they would be expected to if the galaxy's visible, luminous matter represented most of its mass. Roughly an order of magnitude more mass is probably distributed in a large sphere of dark matter, in which the luminous galaxy is embedded. Observational measurements, combined with cosmological arguments, suggest that the mass associated with galaxies, including the mass of the dark matter in which they are embedded, may provide only about a fifth of the total mass density in the universe. The rest could be associated with dark matter (perhaps made up of exotic new kinds of matter) distributed elsewhere.

equivalent to the mass of 10 hydrogen atoms per cubic meter of space.

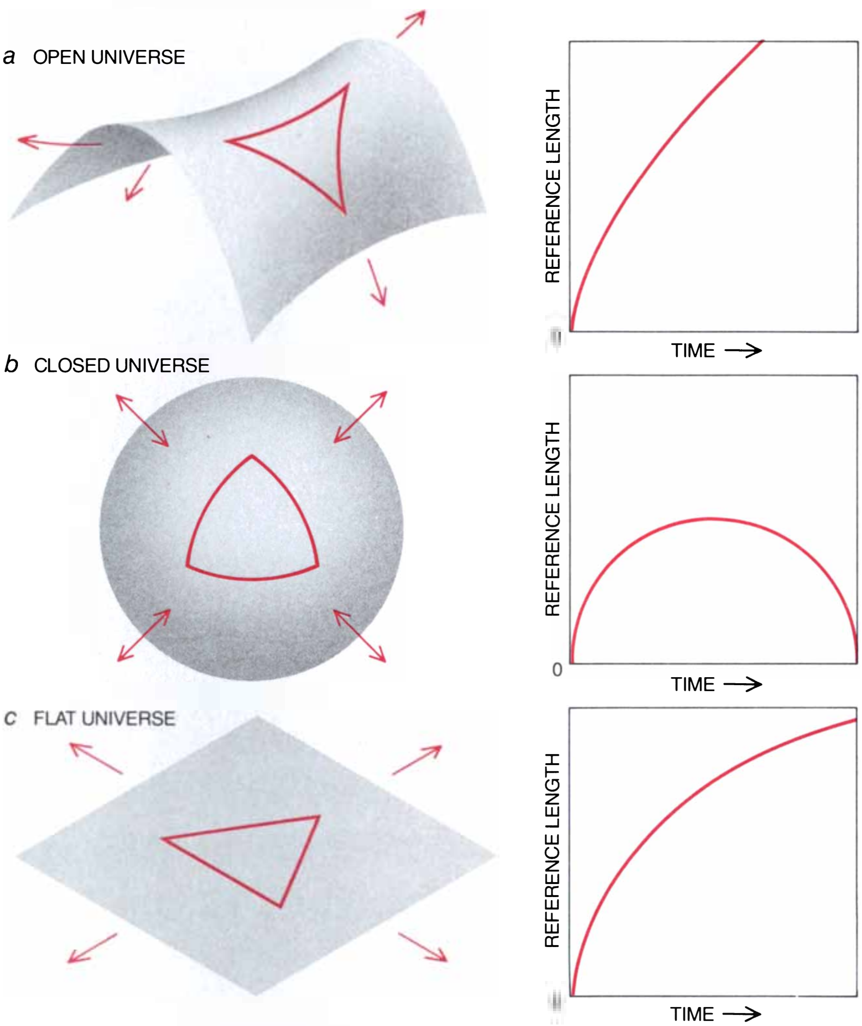
How is it possible to determine how much mass actually exists? One method for finding at least a lower limit is simply to add up the total amount of visible matter. Since what can be measured directly is not mass but luminosity, some amount of interpretation is necessary in translating observations into putative mass densities. When the observed distribution and luminosity of stellar objects and diffuse gas are taken in combination with theoretical estimates of their masses, it seems that the mass-to-luminosity ratio of the luminous matter associated with galaxies is a few times the mass-to-luminosity ratio of the sun. Given this estimate and estimated lower limits on the Hubble constant, the average density of luminous matter in the universe is less than about 2 percent of the density needed to halt the universe's expansion.

It has been known since as early as 1933, however, that clusters of galaxies may contain a significant proportion of nonluminous mass. In that year Fritz Zwicky of the California Institute of Technology was analyzing the individual velocities of galaxies within the Coma cluster. He found many galaxies were moving so quickly that the cluster as a whole should tend to fly apart unless there was more mass to hold it together than the luminous mass alone. Other evidence indicated the cluster was stable, and so Zwicky concluded that the cluster must contain nonluminous matter.

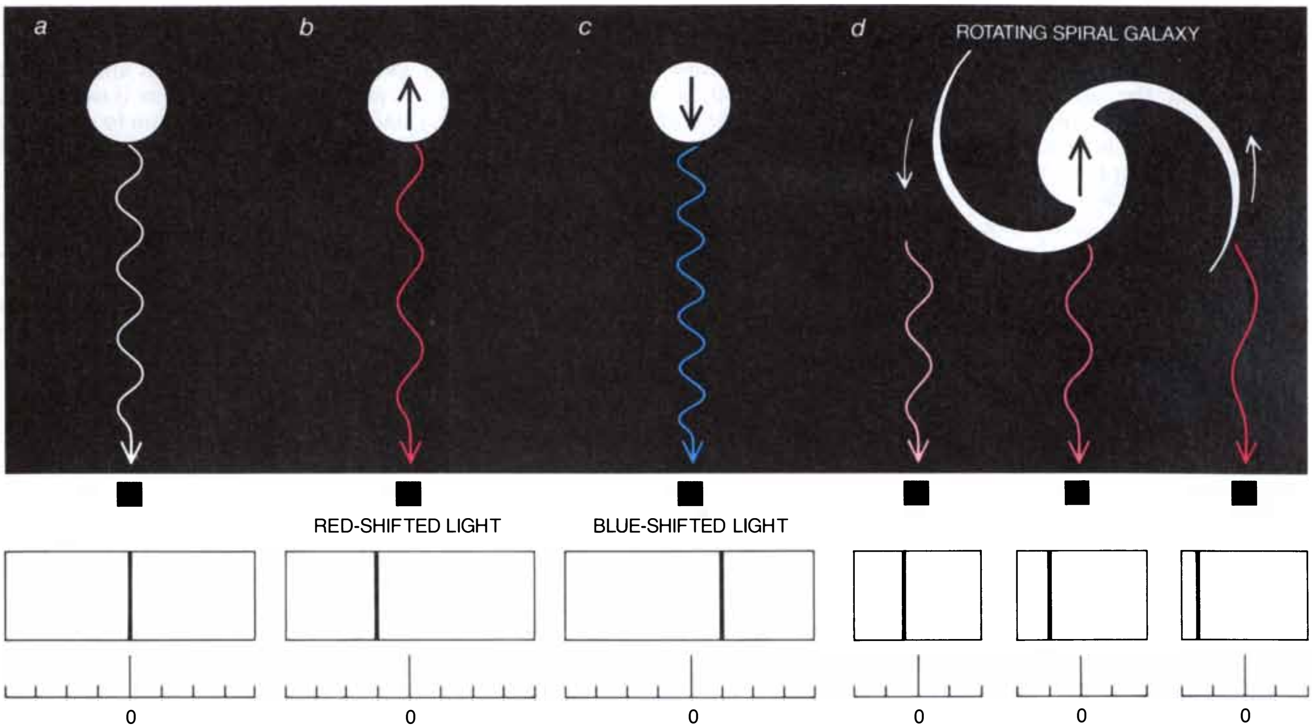
Zwicky set an important precedent by showing that dark matter can in principle be detected indirectly by its gravitational effects. In recent years investigators have shown convincingly that similar techniques can detect the presence of dark matter in structures on scales ranging from the immediate solar neighborhood through galaxies and clusters of galaxies to superclusters made up of thousands of galaxies.

The best-documented evidence for the presence of dark matter is based on the velocities of rotation of spiral galaxies [see "Dark Matter in Spiral Galaxies," by Vera C. Rubin; Scientific American, June, 1983]. The velocity of rotation of an object in a stable, gravitationally bound system, such as a spiral galaxy, depends in part on its distance from the center of rotation. According to Newton's laws, the orbital velocity of objects far from a central concentration of mass should drop off in proportion to the reciprocal of the square root of their distance from the center of rotation. In extensive surveys of stars and hot gas in the outer regions of spiral galaxies, several groups have shown that the rotational velocities of these objects remain constant, rather than dropping off, out to distances greater than 30 kiloparsecs from the galactic core. It had already been suggested by Jeremiah P. Ostriker and P. James E. Peebles of Princeton University that there must be some unseen mass in spiral galaxies, because otherwise gravitational instabilities would cause the galaxies to collapse into bar-shaped formations. The stability of spiral galaxies, as well as the rates of rotation of their outer arms, could be explained if the galaxies were each embedded in a large, roughly spherical distribution of dark matter.

There is other dynamical evidence for dark matter, on scales both larger and smaller than the scale of individual galaxies. The evidence is obtained not from measurements of



CURVATURE OF THE UNIVERSE depends on the average speed with which it is expanding and the average density of matter within it. If the rate of expansion is high in relation to the amount of mass, the universe is said to be open. Such a three-dimensional spatial geometry is analogous to a particular two-dimensional geometry: the area near the center of a saddle (*a, left*). The shape of space affects the shape of geometric objects. For example, in an open universe the sum of the angles in a triangle would be less than 180 degrees, just as it is on the surface of a saddle. The effect would not be noticeable over distances as small as those measurable on the earth, just as a triangle on a very small section of a saddle would appear to be normal. If the universe is open, it will go on expanding at a finite rate forever: any reference length (the distance between any two regions of the expanding universe) will continue to increase (*a, right*). If the amount of mass in the universe is high in relation to the expansion rate, the universe is closed. It closes on itself in much the same way as the surface of a sphere closes on itself (*b, left*). On a sphere the angles of a triangle add up to more than 180 degrees. If the universe is closed, it will eventually stop expanding and will then contract once again (*b, right*). If the recession rate and the amount of mass in the universe are exactly matched, the universe is flat and analogous to a plane (*c, left*). It will continue expanding, but the rate of expansion will slow asymptotically (*c, right*). Strong theoretical arguments support the proposition that the universe is actually flat, even though in order to be flat it would have to contain much more mass than has yet been observed, either directly or indirectly.



DOPPLER FREQUENCY SHIFT makes it possible to determine how quickly a light-emitting object is moving toward or away from an observer and how fast the arms of a spiral galaxy are rotating. A stellar object emits light at characteristic frequencies determined by its composition (*a*). If the object is moving away from the observer (*b*), the wavelength of the observed light appears to be lengthened. This is called a red shift, because longer-wavelength light is redder. If the object is moving toward the observer (*c*), the wavelength of the light is shortened; the light is blue-shifted. Spiral galaxies tend to be moving away from the earth because of the expansion of the universe. To an observer on the earth the light from the center of a spiral galaxy therefore appears to be red-shifted (*d, center*). One arm of the spinning galaxy (*d, left*) will not be moving away from the earth as quickly as the galactic center is, and so its light will be less red-shifted. The other arm will be moving more quickly away from the earth than the galactic center is (*d, right*), and so its light will be even more red-shifted. By comparing the red shifts of the galactic center and the arms one can determine the rate of rotation of any part of either arm. It is then possible to infer the distribution of mass in the galaxy.

rotational velocities but from measurements of the random individual velocities of objects within gravitationally bound systems. A well-known theorem of classical mechanics called the virial theorem establishes a relation between the average kinetic and gravitational potential energies of objects in stable, gravitationally bound systems that have reached dynamical equilibrium. It should therefore be possible to estimate the total mass of such a system (which is related to its total gravitational potential energy) by measuring the relative velocities of a large number of pairs of objects within the system. This method has yielded evidence of dark matter in a wide variety of systems, ranging from dwarf spheroidal galaxies as small as 10^7 solar masses to clusters of galaxies as large as 10^{15} solar masses. On the largest scales probed by this kind of analysis (regions within roughly a megaparsec of galaxies) the average mass densities are no larger than about 20 percent of the density needed to close the universe.

Another method, pioneered by Peebles and his co-workers, relies on statistical analysis of large numbers of galaxies rather than on data taken from individual galaxies or clusters. Peebles showed that by amassing statistical data on galactic motion and clustering on different size scales it is possible, under the assumption that the regions probed contain gravitationally stable dynamical systems, to relate the mean relative velocity of a large number of pairs of galaxies to the mean mass density of the universe.

It is striking that all the available methods, including those I have discussed and several I have not mentioned, yield essentially the same result: if the distribution of galaxies traces the distribution of mass in the universe, then the universe contains less than about 20 to 30 percent of the mean mass density that would be necessary for closure.

Even if galaxies are not good tracers of mass, or if somehow all the analyses have involved systematic errors, there is still good reason to believe that at any rate the total amount of ordinary mass (mass consisting mainly of protons and neutrons) in the universe accounts for no more than about 20 percent of the amount that would be required for closure. The evidence comes for the most part from the theoretical framework that explains the process of nucleosynthesis, in which various cosmically abundant light elements and isotopes were first formed.

Nucleosynthesis of light elements occurred primarily in the first few minutes of the universe's existence. The process of nucleosynthesis would have been extremely sensitive to the absolute density of protons and neutrons at that time. In order for the predictions of current theoretical models of nucleosynthesis to agree with the present-day abundances of the light elements, the total density of protons and neutrons that could have been present at the time of nucleosynthesis is constrained so tightly that these particles' current density must be less than about 20 percent of the density required for closure. Thus it seems that if the universe is closed, at least 80 percent of the total mass in it is made up of some other kind of matter.

Since such fundamental theoretical arguments limit the amount of normal mass in the universe to 20 percent of the critical density, and since observational evidence suggests that the mass density associated with galaxies and clusters of galaxies is about that amount, why should cosmologists not assume the universe is in fact open? It is by no means impossible to imagine a form in which enough normal matter to explain the dynamics of galaxies and clusters could remain unseen. Why, then, is there a need to postulate any other form of mass? Why is there a larger dark-matter problem?

Two theoretical barriers stand in the way of the simple assumption that most or all of the mass in the universe is composed of normal matter and that the mean density is only 20 percent of the critical amount. The first barrier is set by a combination of the theory of galaxy formation and observations of the background of microwave radiation that pervades the cosmos.

It is generally assumed that galaxies eventually formed when regions of the early universe that were denser than the average condensed under the force of gravity until they separated from the background expansion to form isolated bound systems. For roughly 100,000 years after the big bang, ordinary matter could not condense in this way. Ordinary matter was still too hot for its constituent particles to have combined into electrically neutral atoms, and so it consisted of independent charged particles. Because ordinary matter was ionized in this way, its microscopic motion was strongly influenced by background fields of electromagnetic radiation: matter and radiation were coupled. Regions of ordinary matter that were denser than surrounding regions and smaller than the horizon size (the distance a light ray could have traveled since the big bang, and therefore the maximum distance over which physical systems could be in causal contact) could not have condensed further, because the "pressure" of the radiation combated the attracting force of gravity.

Eventually the universe had cooled enough for oppositely charged particles to combine, rendering normal matter electrically neutral, and so matter decoupled from radiation. The thermal background-radiation bath to which the matter had been coupled was then free to cool as the universe expanded, and it now constitutes the well-known cosmic microwave background radiation, which fills the universe. Observations have shown that this background radiation is isotropic—the same in all directions—to within a very high degree of accuracy.

Since gravity is a universally attractive force, any initial fluctuations, or small variations, in the density of ordinary matter in the early universe would have tended to grow after the force of radiation pressure no longer acted against the force of gravity. Thus it is presumed that the universe became (and is becoming) clumpier with time and that galaxies, whose cores now have densities more than one million times the average background density, began in fluctuations whose densities were much closer to the background value.

How large were the initial fluctuations? Because of the limited data currently available on large-scale structures, and because of the mathematical difficulties inherent in describing analytically the evolution of systems as dense as galaxies, it is extremely difficult to work backward from the current state of the universe to determine the precise nature of the initial fluctuations. An easier approach is to assume some initial pattern of fluctuations, simulate the growth and evolution of that pattern and compare the result with present-day observations. In this approach the cosmologist is guided by both lower and upper limits on the size and nature of the initial fluctuations. First, they must have been extreme enough (that is, the ratio between the local overdensity in the region of the fluctuation and the average density in space must have been large enough) for fluctuations on the scale corresponding to galactic sizes to have condensed to form galaxies by today. Second, the fluctuations must have been of small enough amplitude for them not to have left an anisotropy in the background radiation larger than the measured upper limit.

These two conditions appear to be mutually inconsistent if the universe is composed mainly of normal matter. Between the time when normal matter became decoupled from radiation and the time when the fluctuations that would become galaxies collapsed to form isolated, gravitationally bound systems, the initially small fluctuations in density could grow only at a well-defined rate. Fluctuations large enough to have had sufficient time to form self-bound systems would have led to an anisotropy in the background radiation more than an order of magnitude greater than the observational upper bounds. In other words, there has not been enough time, since decoupling, for galaxies to form gravitationally from variations in density small enough not to have left observable traces in the background radiation.

This conclusion depends on two widely held assumptions, namely that the microwave background has not been significantly disturbed since the time of decoupling and that gravity alone led to the formation of galaxies. Unless either of these standard assumptions is false (as various investigators have suggested), it appears that some new form of matter is necessary, one that could have begun to condense gravitationally earlier than normal matter could have.

There is a second and more fundamental reason to suppose the universe is not dominated by normal matter having a density of only about 20 percent of the critical density. This reason, now called the flatness problem, was first pointed out by R. H. Dicke of Princeton and Peebles. The essential point is that any deviation from an exactly flat universe should tend to increase linearly with time. If the universe had had even a small nonzero curvature at the time of nucleosynthesis, the deviation from flatness would by today have increased by a factor of about 10^{12}. Since the mass density in the present-day universe is within a factor of 10 of the mass density of a closed universe (in other words, since the universe is relatively close to being flat), at nucleosynthesis the universe must have been either exactly flat or curved to an extremely small degree: it must have been flat to an accuracy of within one part in a million million.

If the universe is measurably curved today, cosmologists must accept the miraculous fact that this is so for the first time in the 10^{10}-year history of the universe; if it had been measurably nonflat at much earlier times, it would be much more obviously curved today than it is. This line of reasoning suggests that the observable universe is essentially exactly flat: that it contains precisely the critical density of mass. Since normal matter probably accounts for only 20 to 30 percent of the critical density, some form of more exotic matter is probably present.

The next logical question is: Why is the universe exactly flat? In 1980 Alan H. Guth, now at the Massachusetts Institute of Technology, proposed an answer. It took the form of a model of the evolution of the early universe based on ideas in particle physics that had only recently been proposed.

Guth drew on the work of Howard Georgi and Sheldon Lee Glashow of Harvard University. In 1974 the two investigators proposed that three of the fundamental forces of nature—the so-called strong, weak and electromagnetic forces—are different aspects of a single, "unified" force. At sufficiently high energies the three forces should be exactly symmetrical: they should behave identically. At energies comparable to those observed now on the earth, on the other hand, the three forces can behave quite differently [see "A Unified Theory of Elementary Particles and Forces," by Howard Georgi; SCIENTIFIC AMERICAN, April, 1981]. The temperature of the early universe, soon after the big bang, was initially high enough for the symmetry of the three forces to be manifest. As the universe cooled below the critical energy at which the symmetries relating the forces can be maintained, the preferred configuration of the universe became one in which the symmetry was "broken." The effect of this symmetry breaking was that the forces appeared distinct from one another.

(A simple example of this type of behavior is found in ferromagnets. At sufficiently high temperatures a piece of iron is not magnetized: the spins of all the electrons, each of which causes a small magnetic field, point in random, different directions. Below a certain critical temperature, however, it may be energetically more favorable for all the spins to point in one direction, aligning their magnetic fields and creating a permanent magnet. The direction of the magnetic field in the magnet represents a unique direction, and so the symmetry of the former configuration, in which no direction was special, is broken.)

According to Guth's idea, which was later extended by Andrei D. Linde of the P. N. Lebedev Physical Institute in Moscow and by Paul J. Steinhardt and Andreas Albrecht of the University of Pennsylvania, the abrupt breaking of symmetry could have caused the universe to "inflate" rapidly: the universe could have expanded exponentially, growing by more than 28 orders of magnitude in less than 10^{-30} second. After the period of rapid inflation the universe could have reverted to its normal, nonexponential expansion, which is observed today [see "The Inflationary Universe," by Alan H. Guth and Paul J. Steinhardt; SCIENTIFIC AMERICAN, May, 1984].

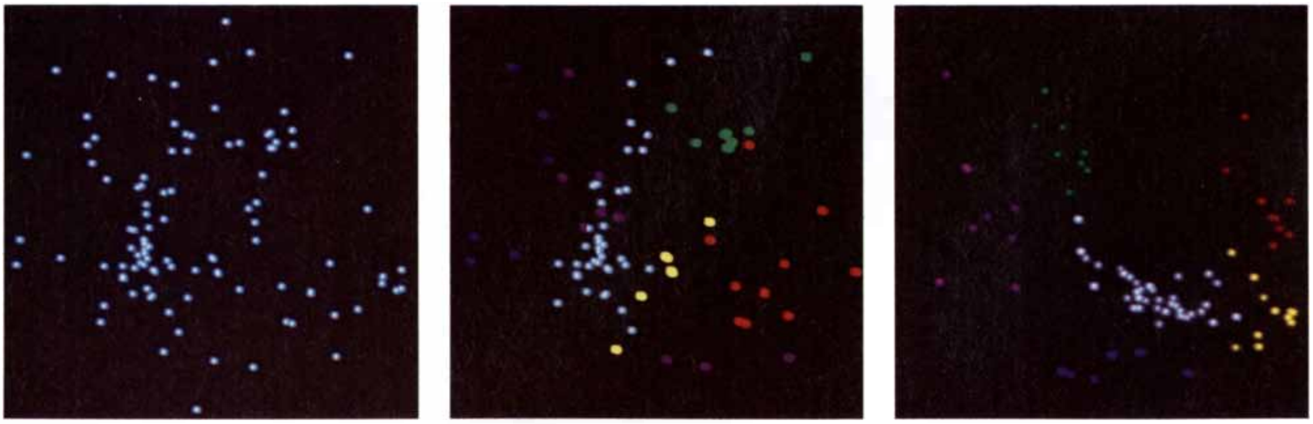
It is the rapid inflation of the universe, according to this model, that caused the observable regions of space to become flat, in much the same way as inflating a balloon makes its surface appear flatter; after inflation the part of the universe observed today would necessarily appear flat.

In addition to its resolution of the flatness problem, the inflationary-universe scenario is remarkably successful in other ways. In particular, it is the only model consistently tying the initial conditions that caused the universe's expansion to the laws of microphysics. The inflationary model also makes it possible to calculate, from first principles, quantities whose values had previously been assumed or inferred. For example, the model remarkably predicts the shape of the spectrum of primordial density fluctuations (the functional relation between the amplitude of fluctuations and their scale size) to be precisely the shape that had been suggested earlier on phenomenological grounds. The wide acceptance by many cosmologists of the predictions of the inflationary-universe model indicates the deep impact particle theory is having on modern cosmology.

In solving the flatness problem, the inflationary model makes the dark-matter problem more urgent. If the universe is flat, then most of the mass in the universe is probably not normal matter, and most of it has not yet been detected in any way, even indirectly.

What might this exotic, undetected matter be made of? One of the earliest proposals was that the dark matter is composed of neutrinos. First postulated in order to solve problems involving the conservation of energy and momentum in nuclear decay, neutrinos interact very weakly with normal matter and are thus extremely difficult to detect. Nevertheless, three kinds of neutrino, called the electron neutrino, the muon neutrino and the tau neutrino, have now been found experimentally. It was originally proposed that neutrinos were massless, but there is no theoretical reason for supposing they might not have some mass. Stringent experimental limits have nonetheless been set on the maximum possible neutrino mass, and it is very small indeed. The strongest constraint is on the electron neutrino, which must have a mass less than about 10,000 times smaller than the mass of the electron.

As dark-matter candidates, neutrinos have two strong advantages over other contenders. First of all, they are known to exist. Second, the calculations that have been so successful in describing primordial nucleosynthesis also suggest that light neutrinos must be abundant in the universe today. When big-bang nucleosynthesis start-



CLUSTER OF GALAXIES IN CANCER, as these computer-generated views show, is not a single dynamical system. As it appears from the earth (*left*) the cluster seems to be a roughly spherical system in apparent equilibrium. Measurements of the velocities of individual galaxies within the system revealed that it would tend to fly apart unless it contained a large amount of unseen matter. Later analysis by Gregory Bothun and his colleagues at the Smithsonian Astrophysical Observatory showed that the cluster is in fact made up of several groups of galaxies separated in space (*center, colors*). A rotated view, in a three-dimensional space in which two axes represent position as seen in the sky and the third represents red shift (*right*), shows the separation of the various groups more clearly. Within each group the relative velocities of galaxies are much lower than the relative velocities of the groups (which are not in dynamic equilibrium), indicating there is less mass in the system as a whole than had previously been estimated. The high relative velocities of the groups themselves had biased the earlier work. This analysis shows that the mass contained in clusters such as the Cancer cluster cannot by itself account for enough mass to produce a flat universe; additional mass must be distributed elsewhere. The computer images here were made by Michael J. Kurtz of the Smithsonian Astrophysical Observatory.

ed, at temperatures greater than 10^{10} degrees Kelvin (degrees Celsius above absolute zero), light neutrinos were kept in thermal equilibrium with matter by the weak interaction and were therefore as abundant as photons. Thus, as R. Cowsik of the Tata Institute of Fundamental Research in India and J. McLelland of the University of Melbourne first estimated, if neutrinos have approximately the same present-day density as the photons that make up the background radiation, and if they have a mass in the range of one ten-thousandth to one hundred-thousandth the mass of the electron, they could account for enough mass to close the universe. (The estimate was later confirmed by more detailed calculations.)

This point became particularly relevant in 1980 when V. A. Lubimov and his collaborators at the Institute of Theoretical and Experimental Physics in Moscow announced they had found evidence that the electron neutrino has a mass within that range. On the basis of this result it seemed neutrinos were ideal candidates to be the dominant mass in the universe. Since then, however, the likelihood that light neutrinos are the dark matter has become much smaller. In the first place, there are many outstanding experimental questions about the Soviet result; as a matter of fact, a recent finding by a group at the Swiss Institute for Nuclear Research appears to contradict it. In addition a great deal of work by astrophysicists has shown that theoretical pictures of a universe dominated by light neutrinos are not as compatible with observation as it once seemed.

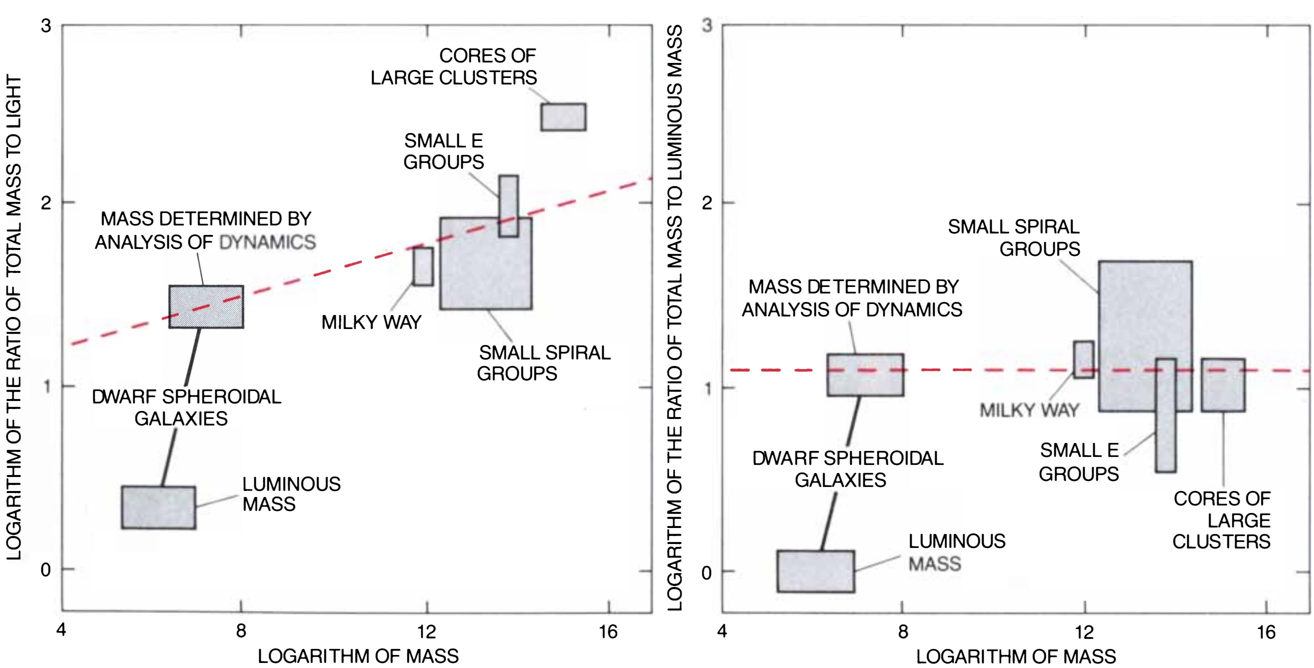
The first such theoretical evidence came in 1979 from investigations by Scott D. Tremaine and James E. Gunn, then both at Caltech. They noted that, for reasons based partly on the Pauli exclusion principle, neutrinos in the relevant mass range could not condense sufficiently to be dark matter on scales much smaller than galaxies. The existence of dark matter on such scales has since been demonstrated convincingly by observations of dwarf spheroidal galaxies.

This work does not preclude the possibility that neutrinos are the dark matter on larger scales. Nevertheless, such a proposal seems incompatible with substantial recent theoretical work describing the evolution of the early universe, which has demonstrated that the large-scale gravitational clustering (the clustering of galaxies and of clusters of galaxies) likely to occur in a neutrino-dominated universe does not seem to resemble the clustering actually observed.

In a neutrino-dominated universe the first structures to form would not be on the size scale of galaxies but rather on the scale of clusters of galaxies or even superclusters (clusters of clusters of galaxies). Unlike normal matter, neutrinos in the early universe were not coupled to electromagnetic radiation. Even so, for some time they were not able to clump together appreciably because, being extremely light, they moved at relativistic speeds, and relativistic objects are not bound gravitationally except by very highly condensed objects such as black holes.

As the universe expanded, neutrinos cooled until they slowed down and became nonrelativistic. At the same time, the radiation background continued to cool to mean energies below those of the nonrelativistic neutrinos. Shortly before the time at which normal matter decoupled from electromagnetic radiation, neutrinos having masses in the appropriate range to close the universe would have become nonrelativistic and would have begun to make up the primary component of the energy density of the universe. Analytic calculations show that only after this time could they have clumped together gravitationally. At any earlier times, fluctuations on scales smaller than the horizon would have been broken up because the neutrinos, being relativistic, would not have been bound to dense regions.

Thus the first scale on which fluctu-



STRUCTURES ON DIFFERENT SIZE SCALES have different mass-to-light ratios, but they have about the same ratio of total mass to luminous mass. Because larger structures tend to have higher mass-to-light ratios (*left*), it might appear that the very largest luminous structures could contain enough mass to make the universe flat. The ratio of total mass to luminous mass (a more physically significant ratio) seems to be constant, however (*right*), indicating that larger structures do not have proportionally more mass than smaller ones. The difference between graphs is accounted for largely by the hot gas (which radiates in the X-ray range but not the optical range) found primarily in larger systems. The graphs are based on work done by George R. Blumenthal, Joel R. Primack and Sandra M. Faber of the University of California at Santa Cruz and Martin J. Rees of the University of Cambridge.

ations could have grown in a neutrino-dominated universe is the scale of the horizon distance at the time when neutrinos could begin clumping gravitationally. This distance scale corresponds to the size of superclusters, not that of galaxies. Soon after it had decoupled, normal matter would have been drawn into the gravitational potential wells caused by clumps of neutrinos. These supercluster-size formations might then have fragmented into galaxies.

That scenario of a neutrino-dominated universe is attractive in many ways. It would have led to a system of filament-shaped superclusters and large "voids" (regions empty of matter) that resemble features identified in current surveys of large-scale clustering [see "Very Large Structures in the Universe," by Jack O. Burns; SCIENTIFIC AMERICAN, July]. In addition, the fact that gravitationally bound formations of neutrinos could begin to grow earlier than systems composed of normal matter indicates that the initial density fluctuations in the universe could have been small enough to be at least marginally consistent with measurements of the background radiation's isotropy.

These attractive features led Carlos S. Frenk of the University of Cambridge, Simon D. M. White of the University of Arizona and Marc Davis of the University of California at Berkeley and, independently, Joan Centrella of Drexel University and Adrian L. Melott of the University of Chicago to develop numerical models investigating the details of gravitational clumping in a neutrino-dominated universe. The investigators encountered serious difficulties when they tried to re-create the clustering that has actually been observed. Essentially they found that in a neutrino-dominated universe the fragmenting of clusters into galaxies and the formation of galaxies would have to have occurred relatively recently (when the universe was at least half its present age) in order to match the currently observed level of clustering. This conclusion is hard to reconcile with the existence of such structures as quasars, which formed in much earlier eras.

In general, the major problem with neutrino-dominated cosmology is that in order for galaxies to have condensed by the present time, structures on much larger scales would have to be much less diffuse than the observed large-scale structures actually are, because structure on the scales of galaxies and superclusters would have formed contemporaneously. Well-defined large-scale clustering would also cause difficulties in matching the predicted random velocities of galaxies in clusters to the observed velocities. For these and other reasons a neutrino-dominated universe now seems implausible.

A way out of the problems with neutrino models seems clear: find models in which galaxies can form significantly earlier than larger structures do. This suggests the need for what has become known as cold dark matter: dark matter that was so cold (that is, moving so slowly) that it was nonrelativistic significantly earlier than neutrinos were and could therefore cluster gravitationally much earlier.

The time at which a class of particles becomes nonrelativistic is a key factor in determining the size of structures that can be formed by that class of particles. At times before the particles become nonrelativistic, structures on scales smaller than the horizon would break up. Hence, in order for galaxies to form before larger structures, cold dark matter would have to have been nonrelativistic by the time the horizon reached the scale size of galaxies.

Ever since the problems with neutrino-dominated theories became clear, a great deal of effort has gone into the analysis of cosmology dominated by cold dark matter, and almost all the results have been positive. Because density fluctuations can grow earlier, the initial fluctuations need not be as large and so any conflict with the observed isotropy of the background radiation is eliminated. Moreover, because cold dark matter could have clumped on smaller scales than neutrinos could have, it might account for the excess mass in such small structures as dwarf galaxies.

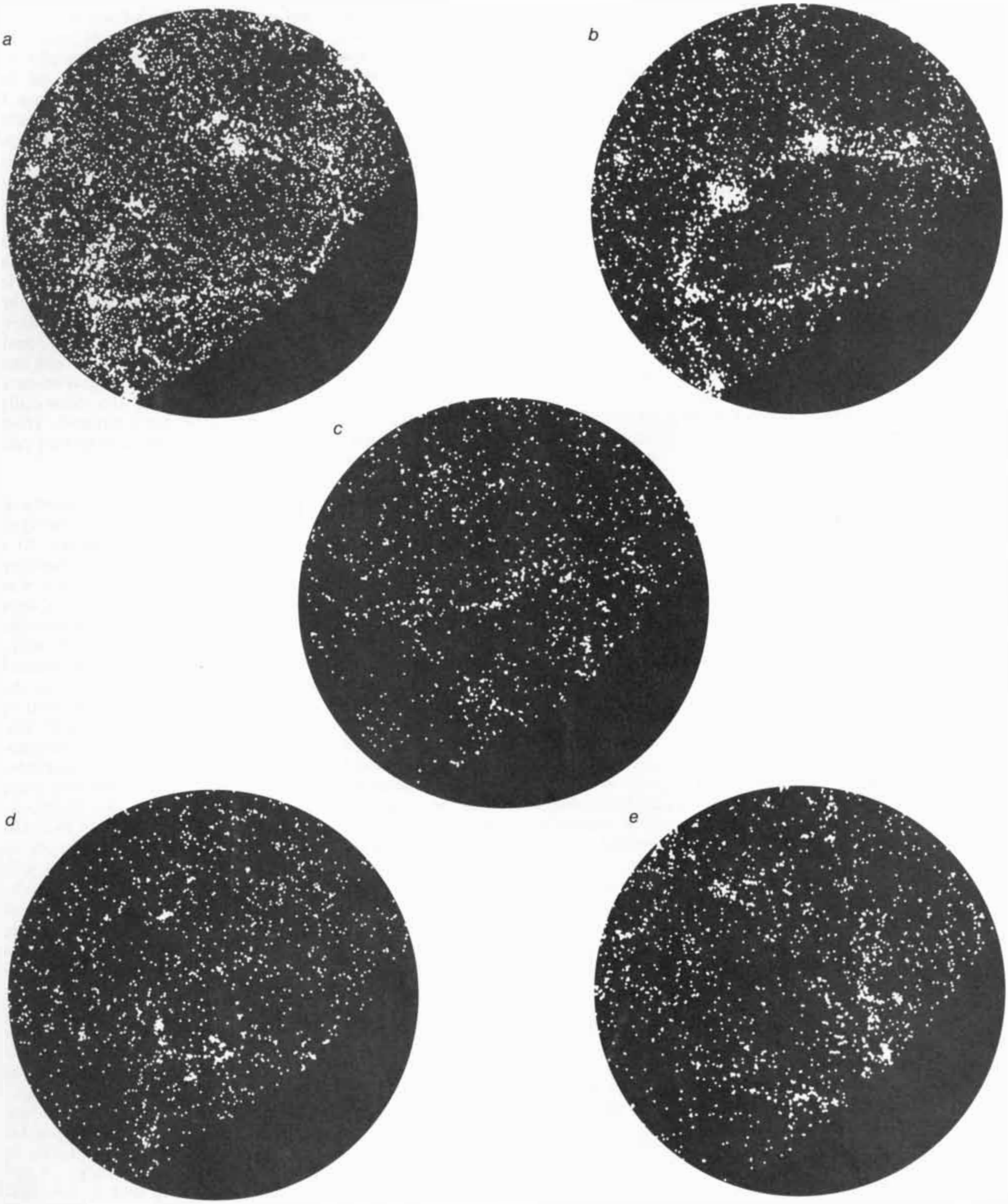
Detailed analytical and numerical investigations are most encouraging. For example, it has been shown that the presence of cold dark matter in the early universe could account in detail for the shape and structure of many types of galaxies. More generally, Frenk and George Efstathiou of Cambridge, along with Davis and White, have shown numerically that clustering on large scales in a universe dominated by cold dark matter can match well with most of the observed features of the actual clustering.

There is still at least one obstacle that apparently prevents complete agreement between theory and observation if the universe is exactly closed and dominated by cold dark matter: Where is the matter? Apparently it can cluster readily on galactic scales, but, as I have described, there is no evidence for a critical density on such scales. One solution to the problem is to assume that galaxies themselves are not good indicators of where most of the high concentrations of mass are: that much of the cold dark matter lies in regions uncorrelated with the locations of these luminous systems. It could well be that galaxies represent statistically rare events, and that most of the mass in the universe has not ever condensed to form galaxies. Examining the clustering of galaxies would then give a biased value for the actual mass density of the universe. The implications of this proposal have been studied in detail, and it appears to lead to scenarios that agree well with most aspects of the observed clustering (with some notable exceptions). Moreover, current work by Frenk and his collaborators suggests that scenarios in which galaxies are statistically rare might arise more naturally from gravitational clustering than had previously been supposed.

The cold-dark-matter hypothesis has forged a strong link between particle physics and cosmology. At a time when cosmologists were deciding some form of cold dark matter was necessary, high-energy physicists were independently proposing the possible existence of new, exotic particles within the framework of various unified theories. As it happens, several of the particles proposed to fill theoretical gaps in high-energy physics could also serve quite naturally as the cosmologists' cold dark matter. These particles have the disadvantage that they have not been observed; unlike neutrinos, they are at this point merely theoretical constructs. Nevertheless, they have the virtue that their existence was proposed independently from cosmology: they were suggested as solutions to quite different problems in particle theory, and yet each of them, for entirely different reasons, could act as cold dark matter.

Among the most attractive candidates on the market today are particles called axions. The existence of axions follows naturally from a theoretical approach developed to explain a special relation that links, in the theory of strong interactions between quarks, the two kinds of symmetry known as charge conjugation and parity.

An interaction is said to be symmetrical under charge conjugation if the interaction would "look" the same were every particle to be replaced by its antiparticle (which has the opposite charge). An interaction is symmetrical under parity if it would look the same when mirror-reflected. The interactions governed by the strong nuclear force (the force that binds quarks together to form protons and neutrons) appear to be symmetrical to a very



NUMERICAL SIMULATIONS of scenarios for the formation of large-scale structure in the universe suggest the dark matter probably does not consist of light neutrinos. It may consist of some kind of cold dark matter: particles that were nonrelativistic in early times, which could have clumped together under the influence of gravity earlier than light neutrinos could have. The diagrams show how galaxies could cluster in each scenario, assuming plausible initial conditions; points represent galaxies. In a universe filled with enough light neutrinos to account for all the dark matter (*a*, *b*), clustering on large scales would be more well defined than is indicated by actual observations (*c*). The observations are matched much better in a universe that was dominated by cold dark matter as it formed (*d*, *e*). In the dark-matter scenarios shown, the universe contains only about 20 percent of the mass needed to make it flat. Simulations in which the universe contains enough mass to be flat do not match the observations well if galaxies trace the dominant mass distribution in the universe. If galaxies do not trace the dominant distribution, simulations of cold dark matter agree well with the observed large-scale structure. The simulations were done by Marc Davis of the University of California at Berkeley, George Efstathiou and Carlos S. Frenk of Cambridge and Simon D. M. White of the University of Arizona.

high degree under a special combination of charge conjugation and parity: the interactions look much the same if all the particles are replaced with their antiparticles and the entire interaction is mirror-reflected. Theoretically this special combination of symmetries need not hold true. The equations governing the strong interactions include several terms that could in principle grossly violate the combination of symmetries.

In 1977 Roberto D. Peccei and Helen R. Quinn, then both at Stanford University, suggested a way to explain why the combination of symmetries is obeyed so well. Their solution was to introduce a new kind of symmetry—a relation between the forms of different fundamental forces that is manifest at sufficiently high energies but is broken at low energies. It was later pointed out by Frank Wilczek, now at the University of California at Santa Barbara, and Steven Weinberg of the University of Texas at Austin that the fact that the Peccei-Quinn symmetry breaks indicates the existence of a new, very light particle. The new particle is the axion. Much recent theoretical work has refined the original model and increased the temperature at which the Peccei-Quinn symmetry is expected to be broken. One of the big surprises to result is that, because the existence of axions depends on symmetry breaking, an axion "background field" might form in the universe, much as a background electric field would exist if the universe were not charge-symmetric (that is, if it did not contain equal numbers of positive and negative charges). Although axions are themselves very light, calculations show that the background field as a whole could clump in much the same way as heavier, nonrelativistic particles would, making the background field an ideal candidate for dark matter.

Another candidate for cold dark matter comes from the theoretical framework known as supersymmetry. In the theory of supersymmetry, for every particle now known there exists a "supersymmetric partner": a particle identical in most respects except spin. Such particles have not yet been observed in the laboratory, and so they must have large masses. Simple models suggest that supersymmetric partners could behave, in their interactions with normal matter, much like very heavy neutrinos. The most promising dark-matter candidate of the supersymmetric partners is the supersymmetric partner of the photon, which is called the photino. Calculations done by me and by others have shown that photinos in the mass range of from one to 50 times the mass of the proton could naturally have sufficient cosmic abundance to close the universe today. Although this proposal has generated a great deal of excitement recently, I should note that the models predicting the existence of photinos lead to other cosmological predictions that are hard to reconcile with observations.

A final candidate, related to the hit parade of cold-dark-matter candidates, is not a particle at all. It is a structure called a cosmic string. Cosmic strings are extended topological defects that might have arisen from symmetry breaking in the early universe. They would take the form of long, thin tubes of constant and very great energy density winding through the universe. Much work has gone into showing that cosmic strings could have evolved in such a way that their total energy density would be less than that required to close the universe. Nevertheless, in a universe dominated by cold dark matter and containing strings, the mechanism of galaxy formation, although it is quite different from mechanisms in standard cold-dark-matter models, might still lead to clustering that matched observations.

What makes all these dark-matter candidates so intriguing at present is the prospect that each of them may well be detected, directly or indirectly, in the near future. Experiments are possible that would rule out or, what is more significant, confirm various ones of the hypotheses. A positive result in any of these experiments would yield invaluable information about the evolution of large-scale structure in the universe and about the fundamental structure of matter, and it might provide a unique mechanism for probing the sequence of events that occurred during the first few seconds of the big-bang explosion itself.

Pierre Sikivie of the University of Florida was the first to point out that cosmic axions, although they interact with other matter extremely weakly, might be detected in microwave cavities (cavities in which electromagnetic radiation in microwave frequencies resonates). A background field of axions oscillating together might produce electromagnetic radiation that could in principle be detected in a microwave device. Wilczek, John Moody of the University of California at Santa Barbara, Donald E. Morris of the Lawrence Berkeley Laboratory and I have investigated this detection scheme in detail and have proposed refinements and alternative schemes. The sensitivity necessary to detect cosmic axions appears to be near the limit of modern technology, although the technology itself is improving rapidly.

Heavy dark-matter candidates, such as photinos, might be detected in several ways. Recently I suggested, as several other workers did independently, that heavy dark-matter candidates in the galactic halo could be captured in the cores of the sun and the earth, where they would accumulate. There, as later calculations have shown, they could collide with their antiparticles (which could also be captured) in annihilation reactions that could produce light neutrinos. The light neutrinos might then escape from the sun's or the earth's core and be measured in large underground detectors. The degree to which such a flux of light neutrinos has not yet been observed puts limits on the masses and densities of heavy dark-matter candidates.

Recently it has been pointed out that heavy dark-matter particles might also be detected directly by devices that are sensitive to very small deposits of energy in very large volumes of material. A variety of new detectors of this type have recently been proposed. One device, put forward by Blas Cabrera of Stanford, Wilczek and me, is designed to measure a small increase in the temperature of a large sample of ultracold silicon or of another pure crystalline material. The increase in temperature would occur when sound waves, produced by impinging dark-matter particles, scattered and randomized. Work by Cabrera, Barbara Neuhauser and Jeffrey C. Martoff at Stanford suggests that the sound waves themselves could perhaps be detected directly [*see illustration on next page*].

Even cosmic strings may soon be detectable, either by their direct gravitational effects on the light from distant quasars and the microwave background (concentrations of energy as dense as cosmic strings should create gravitational fields that would bend light appreciably) or indirectly by measurement of the gravity waves or other radiation they should emit as they evolve.

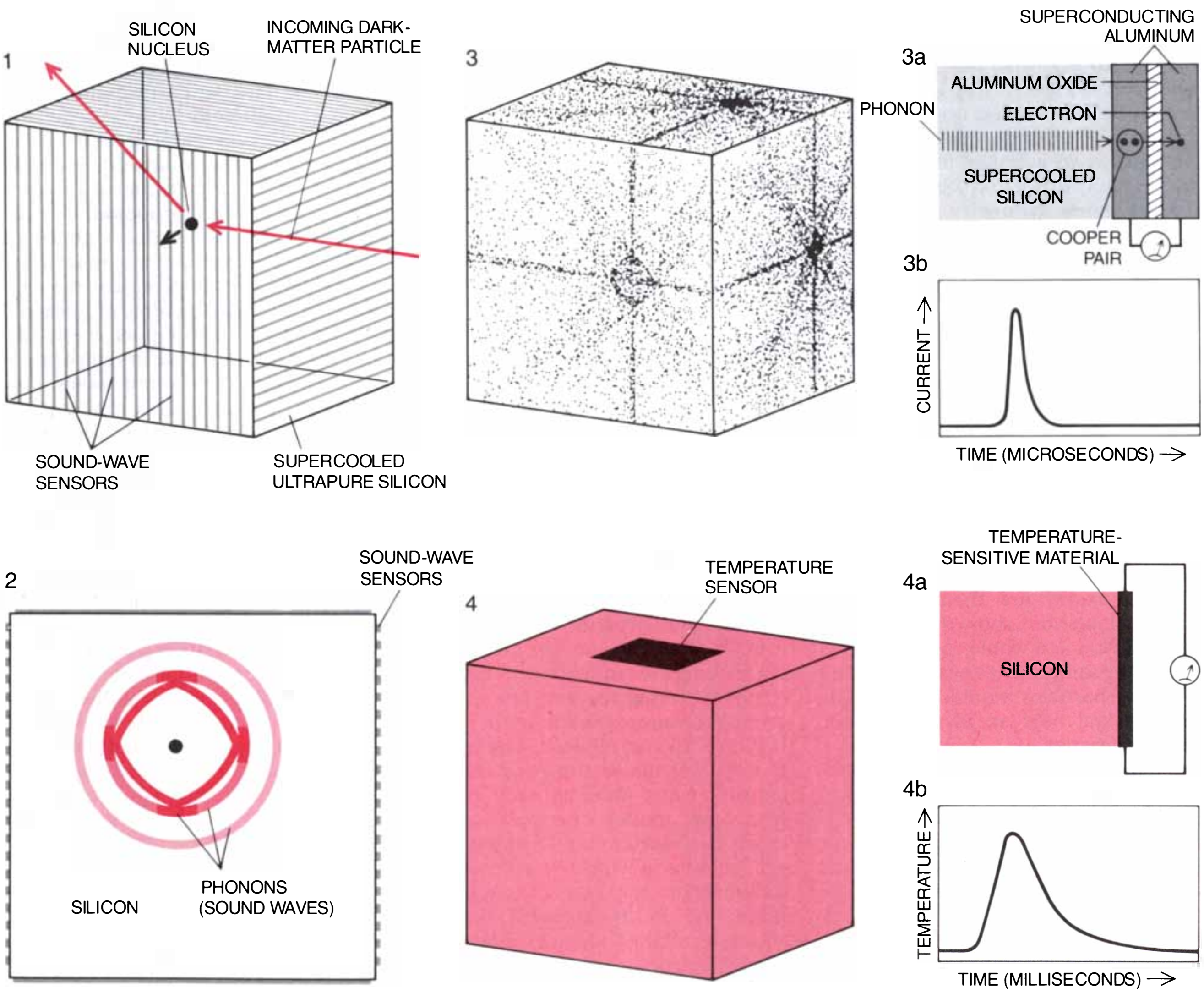
The solution of the dark-matter question could have broad effects on many areas of physics and astronomy. At stake are fundamental notions about both cosmology and particle physics, and it is fitting that each field—often by provoking active debate in the other—has played an important role in the symbiotic evolution of this area of research.

It is important to recognize, however, that cosmology is in many ways in its infancy. There are comparatively few experimental and observational data available for theorists to work with, and so dramatic changes in the field are possible and much of the standard wisdom may be in error. The point is well illustrated by several new results that arose as this article was being written, any of which may have a profound effect on the field.

One new observational result is found in the preliminary analysis of a deep-sky survey being made by Margaret J. Geller, John P. Huchra and their collaborators at the Harvard-Smithsonian Astrophysical Observatory. It seems that nearby galaxies are clustered in filmlike surfaces that surround nearly spherical voids—a structure resembling that of soapsuds or foam bubbles. This remarkable observation, which could completely revise cosmologists' picture of large-scale structure, suggests that forces other than those of gravity are perhaps at work in determining the present-day large-scale structure.

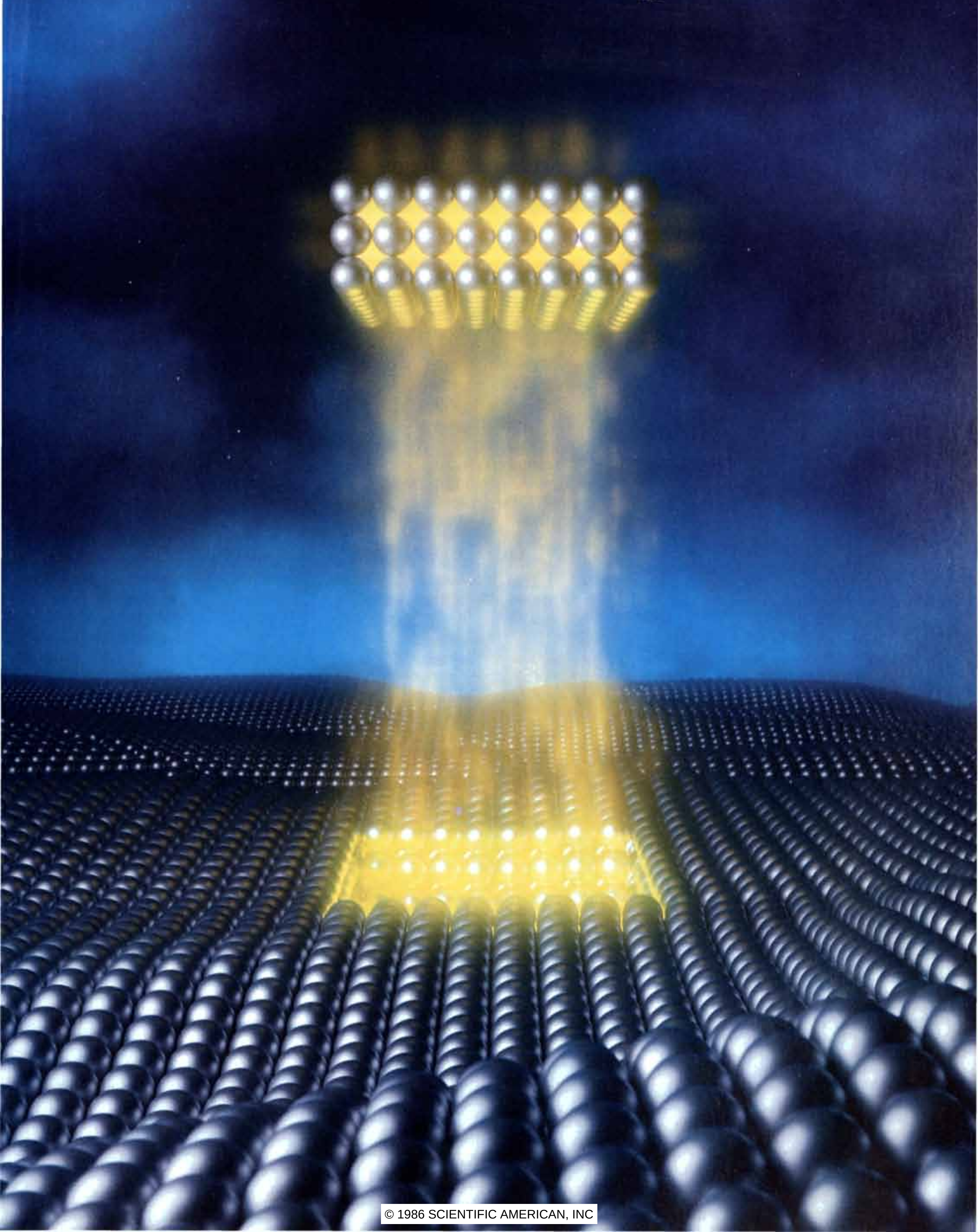
In another new development, work done independently by Tremaine (now at the Canadian Institute of Theoretical Astrophysics) and J. Anthony Tyson of AT&T Bell Laboratories suggests that galactic rotation curves may not be flat indefinitely but rather may drop off at radiuses beyond about 30 kiloparsecs. The work implies that whatever makes up the dark matter may interact more strongly with normal matter than the cold dark matter would be expected to.

Finally, recent data on the motions with respect to the microwave background of very large-scale regions of matter have provided evidence that these regions are moving, together, with an extremely large drift velocity. No current theory of large-scale structure can explain this apparent phenomenon. New measurements such as these, as well as the possibility of detecting the dark matter itself, may soon revolutionize the accepted picture of the universe.



DETECTION DEVICES for dark-matter particles might be made of very pure silicon crystals cooled to within one degree of absolute zero. Such crystals could react measurably to extremely small deposits of energy. In one class of possible detectors, when an impinging dark-matter particle scatters off the nucleus of a silicon atom (*1*), it causes a set of phonons, or sound waves, to spread throughout the material (*2*). Phonons arriving at the silicon's surface will have a distinctive pattern (*3*), which will depend on the location and intensity of the original collision. One detector configuration might detect individual phonons in the pattern as they impinge on the surface of the crystal. To do so the silicon could be overlaid with strips made of two layers of superconducting aluminum sandwiching a layer of aluminum oxide (*3a*). In superconductors electrons are bound together in pairs called Cooper pairs. An incoming phonon might break apart a Cooper pair, and if the aluminum layers are kept at different voltages, the freed electrons might "tunnel" from one aluminum layer to the other, forming an electric current (*3b*). Alternatively, investigators could measure the rise in the temperature of the silicon after the initial energetic phonons had dissipated into a uniform background of random thermal vibrations (*4*). Then the detector could consist of a thin film of a material whose electrical resistance increases sharply with temperature (*4a*). A change in the temperature of the sample as a whole (*4b*) could be determined from the change in resistance.

The Pressure Extrapolation



The Pressure Extrapolation

Modern automotive catalytic converters contain rhodium which promotes chemical reactions to remove pollutants from a car's exhaust. Scientists at the General Motors Research Laboratories have recently made discoveries about one such chemical reaction, the reaction between nitric oxide and carbon monoxide, pointing the way toward new or improved catalysts.

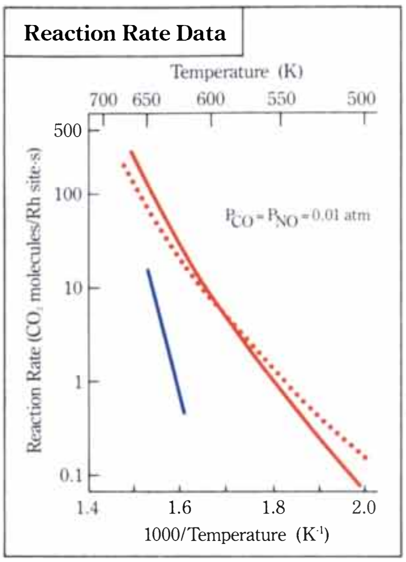


Figure 1: Rate comparisons for the NO-CO reaction. Measured data over single crystal Rh(111) (solid red line) and over supported Rh (blue line); model predictions (dotted red line).

MOST FUNDAMENTAL catalytic studies using surface science techniques require an ultrahigh vacuum environment (10^{-13} atm). They are best suited for studying well characterized materials, such as metal single crystals. Catalytic reactions of practical interest, however, involve polycrystalline materials, in the form of small metal particles dispersed on supports. And they take place at atmospheric pressures rather than in an ultrahigh vacuum.

Now Dr. Galen B. Fisher and Dr. Se H. Oh have demonstrated how the wealth of chemical information obtained from ultrahigh vacuum (UHV) studies of ideal, single-crystal catalysts can be applied to the understanding of real-world systems that have different catalyst environments and that operate at much higher pressures.

These researchers concentrated their studies on the many chemical reactions that occur in modern automotive catalytic converters. One such reaction is the reduction of nitric oxide (NO) by carbon monoxide (CO) over a rhodium (Rh) catalyst to yield carbon dioxide (CO_2) and nitrogen (N_2) (Figure 2).

Dr. Fisher used various surface science spectroscopies in ultrahigh vacuum to study all of the elementary reactions over a rhodium single crystal [Rh(111)] that might be involved in this specific reaction. Over several years he measured the rates and determined the activation energies of each of these reactions. For most of these reactions, this was the first time these parameters had been measured. Based upon these results, Dr. Fisher hypothesized that the elementary reactions shown in Figure 2(a-f) were the significant steps involved in the NO-CO reaction and that nitrogen recombination and desorption (Figure 2f) was the rate-controlling step on Rh(111).

Dr. Fisher and Dr. Oh also initiated kinetic studies of this reaction at realistic reactant partial pressures and temperatures using two different catalysts—one was a rhodium single crystal [Rh(111)], and the other consisted of rhodium particles supported on alumina [Rh/Al_2O_3]. The rhodium concentrations on the support were similar to those used in an automotive catalytic converter. The studies with the single crystal at realistic, high pressures were done in collaboration with Dr. D. Wayne Goodman of Sandia National Laboratories.

At the same time, Dr. Oh devised a mathematical model for this reaction. The model consists

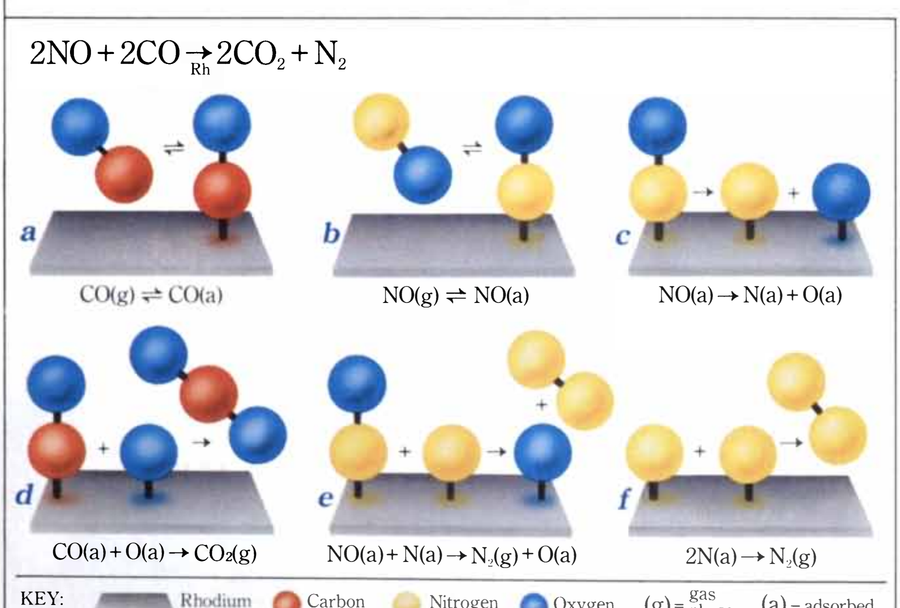


Figure 2: Schematic representation of the elementary intermediate steps for the NO-CO reaction.

of steady-state conservation equations for the surface species, based on the reaction mechanism and the rate expressions for the individual reaction steps determined in Dr. Fisher's UHV studies. Overall reaction rates could then be computed from the surface concentrations satisfying the conservation equations. The reaction rates predicted by this model, which depend only on reactant partial pressures, are shown in Figure 1 (dotted red line).

The kinetics of the NO-CO reaction measured over a rhodium single crystal using realistic reactant partial pressures are shown in Figure 1 (solid red line). The agreement with the model predictions indicates that Drs. Fisher and Oh had correctly identified all of the intermediate reaction steps and confirms that, in this case, nitrogen recombination and desorption (Figure 2f) is the rate-controlling step on Rh(111). The fact that the agreement is so good also indicates that the rates of the elementary reactions measured under UHV conditions are still valid at realistic reactant partial pressures—a pressure extrapolation of more than ten orders of magnitude.

THE KINETICS of the NO-CO reaction measured over the supported rhodium catalyst (Figure 1, blue line), however, were much slower than predicted by the model. In addition, infrared studies have shown that NO is the predominant surface species on the catalyst, suggesting that in this case NO dissociation (Figure 2c) is the rate-controlling step. In fact, if the rate constant for NO dissociation measured under UHV conditions and used in the model is reduced by a factor of 2000, the kinetics of the NO-CO reaction measured over the supported rhodium catalyst are correctly predicted.

The difference between the kinetics of the NO-CO reaction measured over a rhodium single crystal and the kinetics measured over supported rhodium shows that this reaction depends on the environment of the rhodium in the catalyst. The reaction model strongly suggests that the NO dissociation reaction is the reaction step most sensitive to the rhodium environment.

"While our reaction model cannot tell us why NO dissociation is slower on supported rhodium," observes Dr. Oh, "it can help identify the kinds of studies necessary to clarify the origins of such sensitivity." Comparative kinetic studies can also provide useful insights for developing improved NO reduction catalysts. "Our studies have already told us," adds Dr. Fisher, "that one possible path to improving automobile catalysts is to make modifications that increase the NO dissociation rate."

General Motors

GM
MARK OF EXCELLENCE

THE MEN BEHIND THE WORK

Dr. Galen B. Fisher (left) and Dr. Se H. Oh are both Group Leaders in the Physical Chemistry Department at the General Motors Research Laboratories.

Dr. Fisher holds the title of Senior Staff Research Scientist, and heads the Surface Chemistry and Corrosion Science Group. He attended Pomona College as an undergraduate and received his graduate degrees from Stanford University in Applied Physics. Before coming to General Motors in 1978, he did post-doctoral studies at Brown University and worked at the National Bureau of Standards. Since then, his research has been involved with surface science studies of various catalytic reactions.

Dr. Oh is a Senior Staff Research Engineer, heading the Catalytic Kinetics Group. He received his undergraduate degree from Seoul National University and holds a doctorate in Chemical Engineering from the University of Illinois. Dr. Oh did post-doctoral work at the University of Toronto prior to joining GM in 1976. Since then, he has been involved in measuring and modeling the kinetics of catalytic reactions.

Soviet SDIphobia

In the weeks preceding the Reykjavík summit meeting as in the weeks that followed, the leadership of the Soviet Union made it clear that the dramatic reductions in strategic weapons they proposed were intrinsically dependent on restraint of the Strategic Defense Initiative. Why would Mikhail S. Gorbachev and his colleagues establish as a sine qua non concessions in the development of a defense they have publicly declared to be completely ineffective? Does the fear actually lurk in the Kremlin (as SDI proponents suggest) that the program could evolve into an impenetrable shield? Sovietologists and defense experts think there are several more plausible reasons for the Soviet Union to have taken its anti-SDI position, all consistent with the belief that a space-based, virtually leakproof defense against nuclear weapons is unlikely ever to result from the SDI.

First, the Soviets could derive obvious political and propaganda benefits by highlighting their apparent efforts to prevent any new turn in the arms race. Harold Brown, Secretary of Defense in the Carter Administration, sees the Soviet position as an attempt to sow division in Washington and Europe by promoting controversy over trading the SDI bird in the bush for tangible cuts in strategic weaponry.

Soviet spokesmen may also believe that a portrayal of President Reagan's SDI plans as torpedoing Soviet arms-control initiatives could play well on the global stage. Eric Stubbs, an analyst for the Council on Economic Priorities, has pointed out that "the statements associated with [the U.S.S.R.'s] case against SDI occasionally carry the distinct flavor of a public relations campaign."

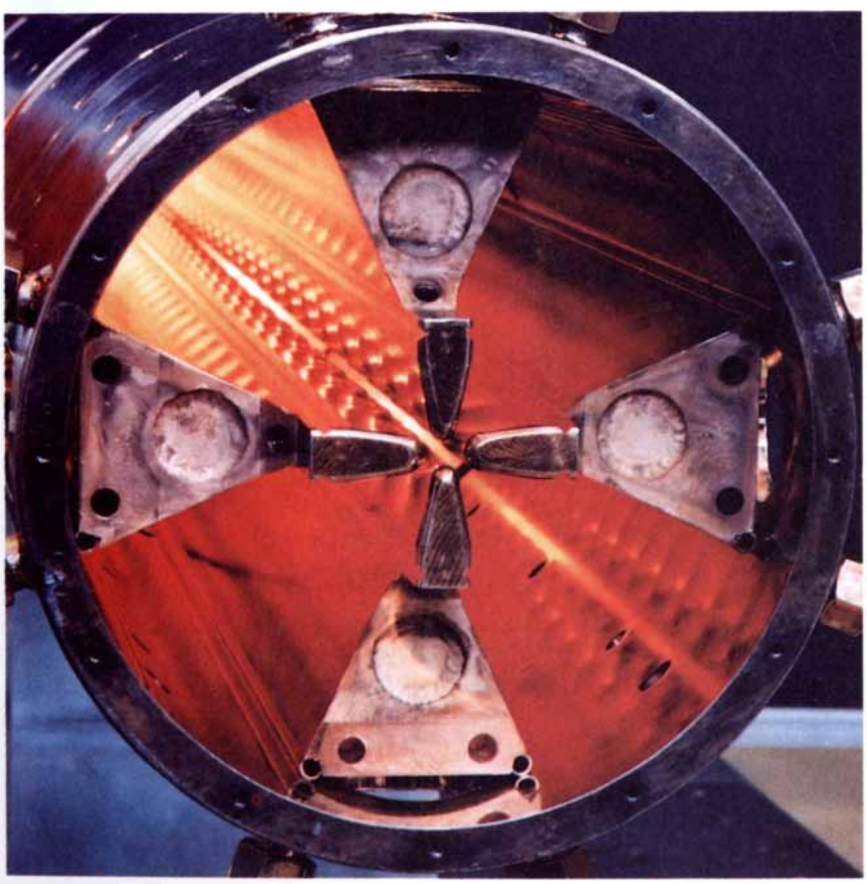
A more material reason, voiced by the Soviets themselves, for opposing the deployment of a strategic defense is that even a mediocre defense can be a militarily useful complement to a first-strike strategic force. An American space shield may be just effective enough to hinder a ragged retaliatory attack—in effect nullifying the Soviet assured-destruction deterrent.

"I think the argument is somewhat overdrawn," Stubbs observes. "It assumes that the U.S. would be willing to take the enormous risks involved in executing a first strike and that the U.S. could, in fact, pull it off; both are questionable assumptions." Brown nonetheless thinks the Soviets "do worry to some degree about this question" because "a strategic defense—whatever its effectiveness—will work better as part of a first strike."

Another important reason for the Soviets' stance, which they do not acknowledge publicly, is the fear that the SDI will create new realms of combat in which U.S. technology would—at least initially—have the edge. Marshall Goldman, associate director of Harvard University's Russian Research Center, says: "The Soviets have enormous respect for U.S. technology—probably more than Americans themselves." In this context it is not the SDI's ostensible goal of an extensive ballistic-missile defense that most worries the Soviets but the application of space-based SDI technologies for offense—directed at satellites in space or even at targets on the ground. As David L. Aaron, former deputy national security adviser, says: "The first fruit SDI is going to yield will be antisatellite weapons. The Soviets are genuinely worried about the near-term advantage the U.S. could gain."

Indeed, the technological applications most feared by the Soviets may be those related to conventional warfare. "They are afraid we would make advances in emerging technologies for conventional uses, where they currently have significant advantages," says Raymond L. Garthoff of the Brookings Institution. Goldman supports such a view by calling attention to the statements made in 1984 by Marshal Nikolai V. Ogarkov (before he was demoted from his position as chief of the Soviet general staff). "He criticized the Soviet military establishment for falling behind in the technology of conventional weaponry. Ogarkov claimed that this posed a greater danger to the Soviet Union than the Western nuclear-weapon threat."

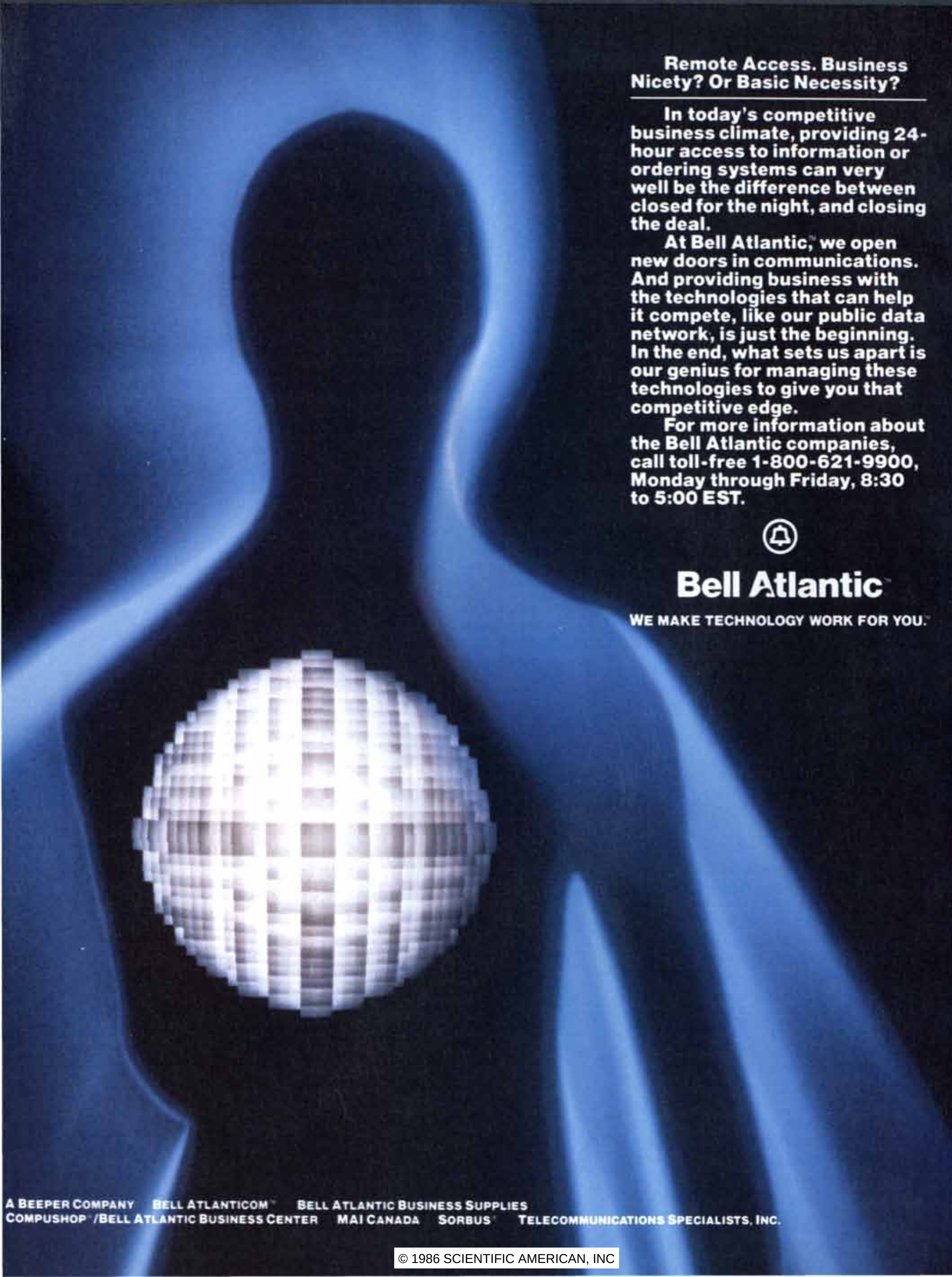
Brown posits another military advantage the Soviets are averse to losing: "They realize that SDI could yield a system that protects the hardened silos of ICBM's. The Soviets can currently destroy our fixed, land-based missiles,

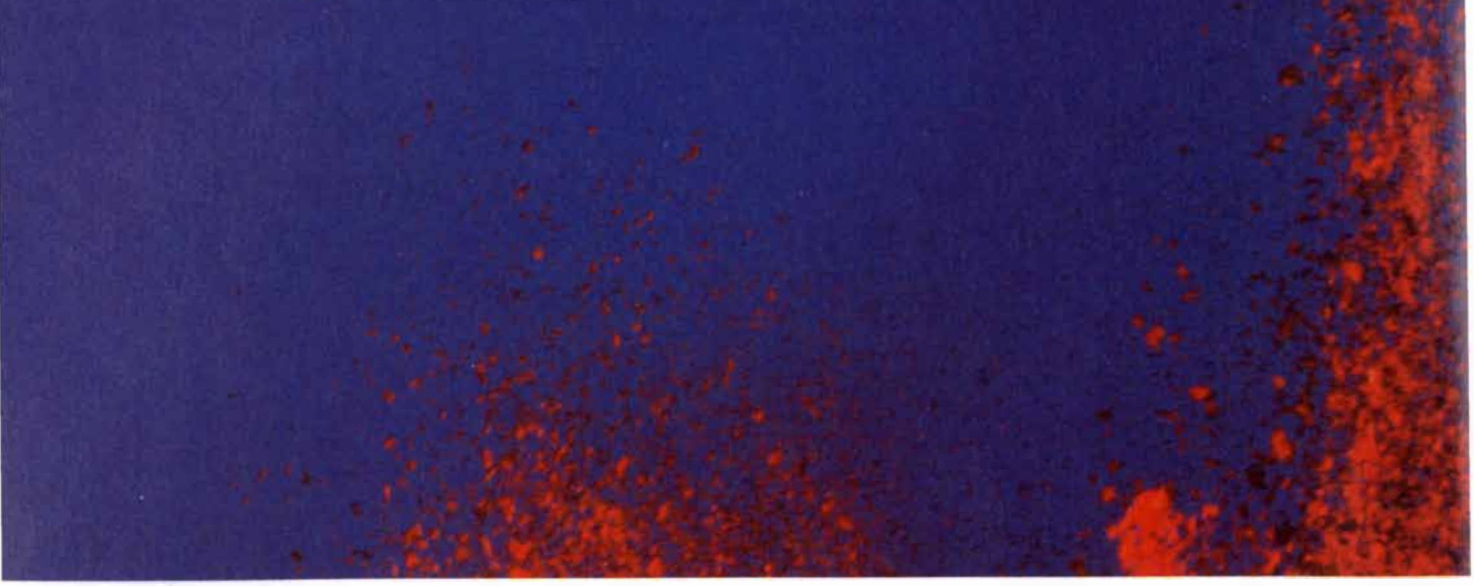


ION ACCELERATOR, developed at the Los Alamos National Laboratory, could be a component of a neutral-particle-beam device. Although such devices are studied under the Strategic Defense Initiative, whose stated purpose is to determine the feasibility of a defensive system, the U.S.S.R. may fear that they could also serve as offensive weapons.

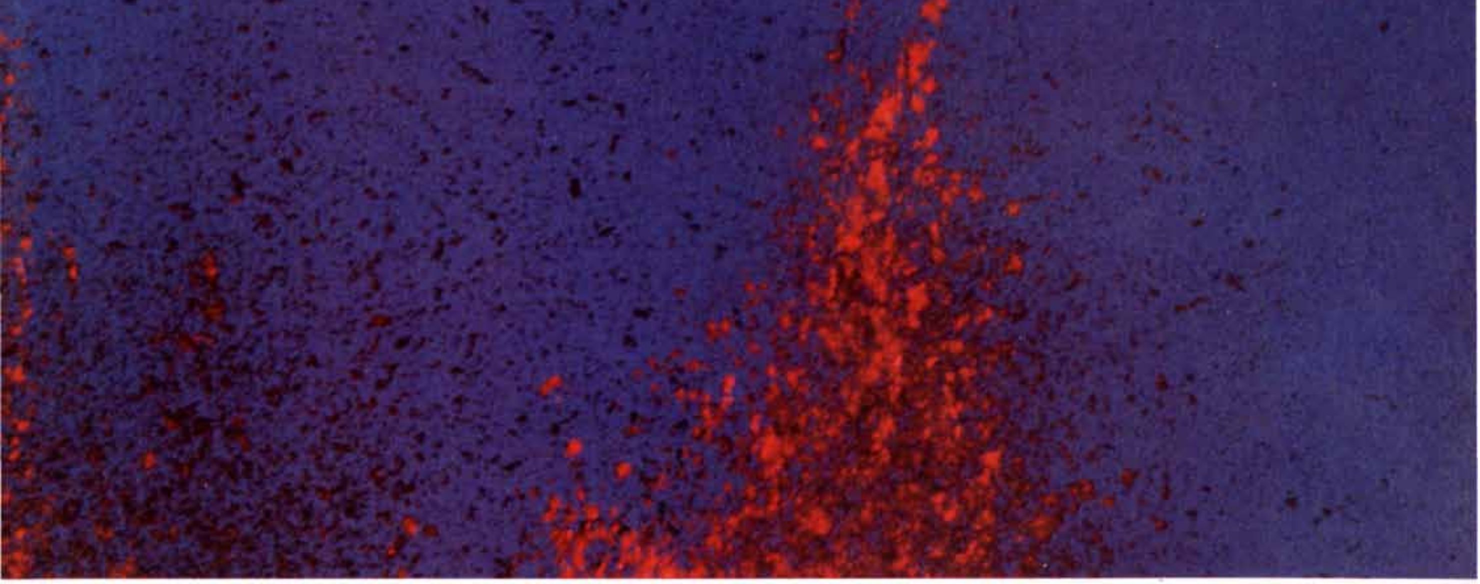

Grand Enchantment.
PRODUCT OF FRANCE
BOTTLED IN FRANCE
Triple Orange
Triple Orange
Grand Marnier
Liqueur
PARIS — FRANCE
750 ml (25.4 fl.oz) · 80° Proof
For gift delivery anywhere call 1-800-CHEER-UP (except where prohibited by law).
Product of France. Made with fine cognac brandy 80 proof. ©1985 Carillon Importers, Ltd., Teaneck, N.J.

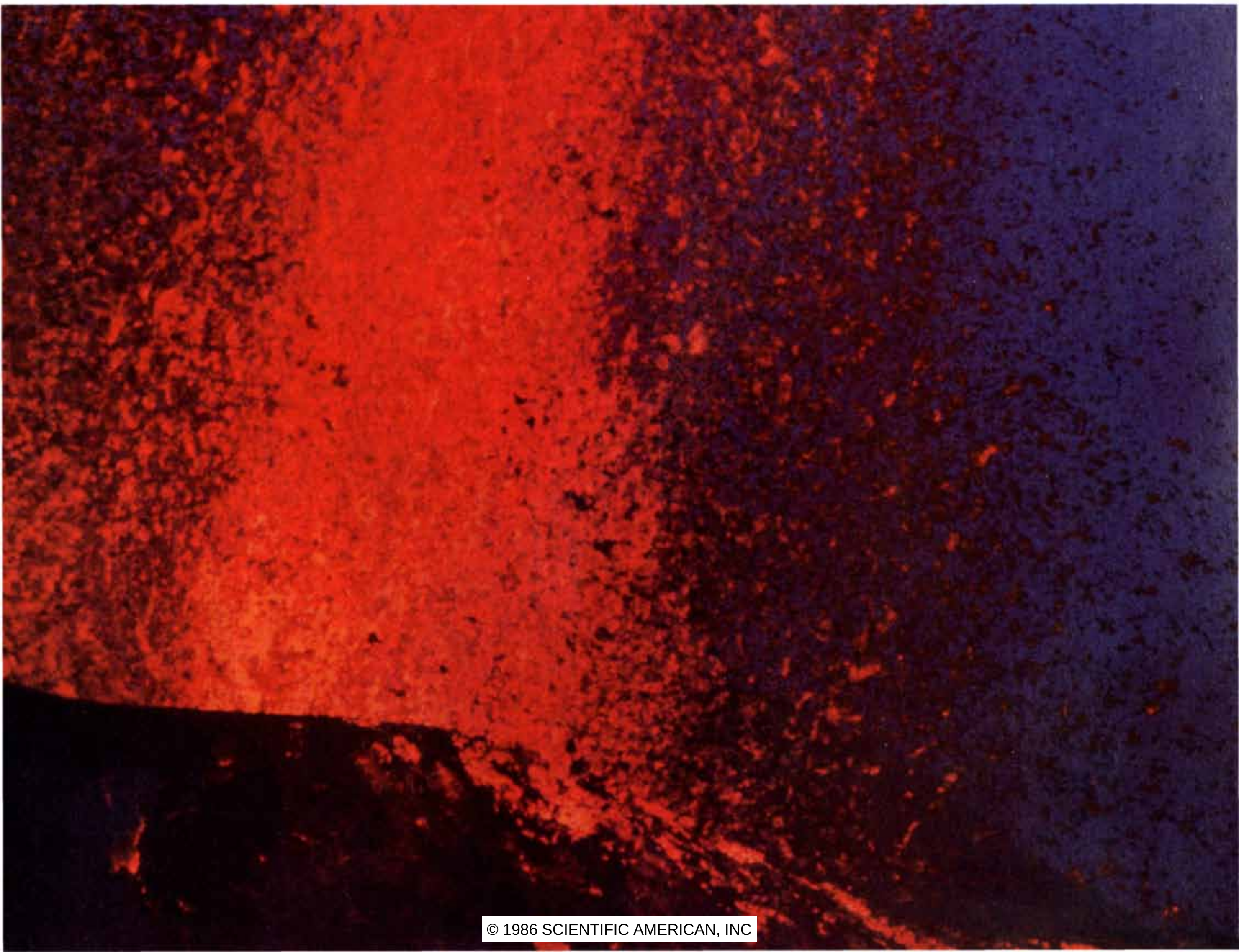
IN TECHNOLOGY ALONE, BUT IN THE ABILITY TO MANAGE IT.






© 1986 Allied-Signal Inc., PO Box 2245R, Morristown, NJ 07960.





and that is an edge they would like to keep. One way to keep that relative edge is to avoid any defense of ICBM's."

Finally and probably most important are the direct and indirect costs the U.S.S.R. would have to pay in responding—through either emulation or countermeasures—to an American deployment of a strategic defense. "The U.S.S.R. could not allow the U.S. to proceed unilaterally in the research and development of such a system," Brown points out. Aaron sees the Soviet Union setting up its own emulative SDI program "because they would want to demonstrate to the world—as they did with Sputnik—that their technology is capable of such things." In any event, Brown says, "it would be prudent for the Soviets to mount a comparable R&D program simply to guard against any technological breakthrough that just might come out of the SDI program."

"No doubt, in addition to whatever else they are doing," observes Brown, "they will increase their offensive forces." Such a move would be an effective countermeasure since an expansion of the number of nuclear warheads and missiles would overwhelm whatever defense the U.S. may place in space. Although Stubbs has calculated that a doubling of the number of ICBM warheads could be accomplished for less than a tenth of the cost of deploying a space-based defense system, "it still is less costly to prevent SDI in the first place."

Indeed, the diversion of scarce technological resources from civilian to military industry in order to engage in such a round of military competition, according to Goldman, would mean that the U.S.S.R. would have to "abort its effort to revamp its economy." Nevertheless, there is almost universal acceptance among observers of the situation that the Soviet Union would take up the challenge if the SDI gauntlet is thrown down by the U.S.

Next Stop: Mars

A consensus seems to be growing within the community of planetary scientists that the exploration of Mars should be made a focus of the nation's space program. Some powerful voices from within this part of the science establishment are already urging that humans go along for the ride.

A number of factors make Mars an attractive goal. Mars is an earthlike planet where huge mountains formed (Olympus Mons at 90,000 feet is some 61,000 feet higher than Mount Everest) and volcanoes were once active. An understanding of these phenomena would shed light on the geologic history of the earth. Mars was once wet and warm rather than cold and dry. It has daily and seasonal weather changes. Apart from satisfying intrinsic curiosity, an understanding of these phenomena could result in better climate and weather forecasting for the earth. An additional spur is that the U.S.S.R. plans to explore Mars, beginning with an unmanned orbital reconnaissance in 1989.

But why send people? Bruce C. Murray of the California Institute of Technology argues that manned missions "reap enormous political rewards: domestic popularity and a unique kind of international prestige." He does not see the *Challenger* disaster as fatal to the domestic popularity of manned space exploration because the shuttle program lost its focus and wound up "putting man in a loop where he wasn't appropriate."

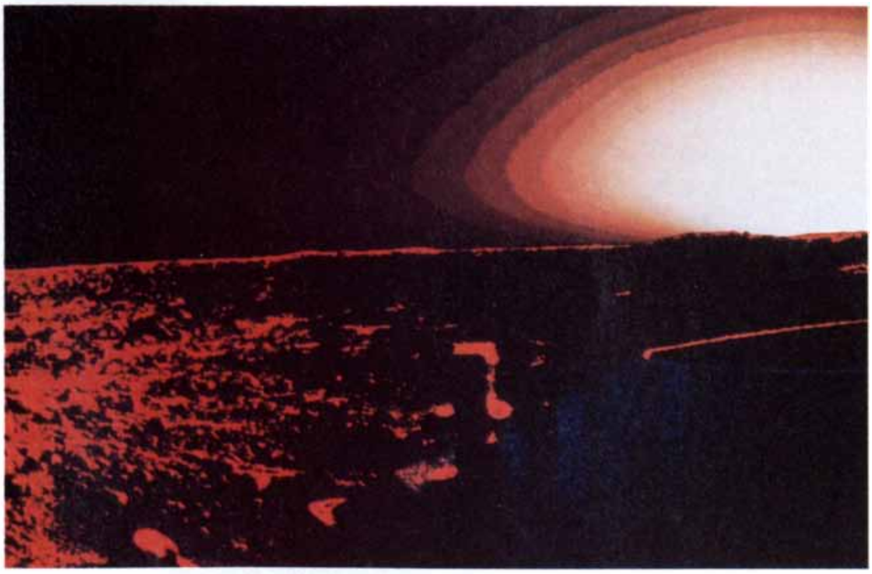
Gerald J. Wasserburg of Caltech argues for unmanned exploration by "smart" roving vehicles that could obtain samples of the Martian surface and reinvestigate areas shown by the samples to be of particular interest. He envisions a cooperative effort with the U.S.S.R. that would also explore and return samples from Venus and a comet. In his view manned exploration of Mars should be a long-term goal for which a sound scientific footing would be laid by the unmanned explorations.

Carl Sagan of Cornell University, who is president of the Planetary Society, also proposes that unmanned "smart" roving vehicles explore Mars for several years. His program would culminate in a joint manned mission to the planet by the U.S. and the U.S.S.R. Sagan says manned exploration is "hugely more expensive" than exploration by semiautonomous rovers and can be justified only by "a reason beyond science." For Sagan, Murray and others the ulterior motive is to take a major step toward world peace.

S. Fred Singer of the University of Virginia (on leave at George Mason University) and several other space scientists propose the establishment of a manned station on the Martian satellite Deimos. From Deimos the astronauts would dispatch dozens of unmanned roving vehicles to Mars to explore its planetary and climatic history, to search for evidence of past or present life and to return samples to a laboratory on Deimos. Singer, who states the case for the venture in *Policy Review,* asserts that deciding on the project now would "excite the popular imagination" and focus the activities of the space program for the next 10 or 20 years.

One element of the reasoning underlying the proposal is that because of Deimos' relatively weak gravitational field a manned mission to that moon would be cheaper than a manned mission to Mars and safer for the astronauts. Another element is that the mission would yield more information about the planet than a few unmanned rovers directed from the earth. Planetary rovers from the earth would have to be complex, heavy and limited in the number of places they could visit and revisit; samples would have to be sent all the way back to the earth.

It remains to be seen what will come of the various proposals. The National Commission on Space, established by



SUNSET ON MARS was photographed in 1976 by *Viking 1.* Further exploration of the planet is being urged by many planetary scientists. Some propose manned missions; others favor "smart" roving vehicles. (NASA photograph from Photo Researchers, Inc.)

SCIENCE/SCOPE®

Advanced spacecraft and military electronics will be made more effective by cooling devices called heat pipes. These devices, which have no moving parts, can improve the performance, reliability, life and cost-effectiveness of computers, signal processors, communications devices, and other equipment. A heat pipe basically is a sealed container with a small amount of fluid (usually water, methanol, or ammonia) and a capillary-wick structure inside. In operation, the fluid vaporizes within the heat pipe at its hot end. The vapor travels to the cool end where it condenses and releases its heat. Then, the capillary-wick structure serves as a pump that transfers the condensed liquid back to the hot end of the heat pipe to begin the cycle again. The U.S. Navy has awarded a contract to Hughes Aircraft Company to develop manufacturing technology for heat pipe assemblies for printed wiring boards.

A new launcher for AMRAAM and Sidewinder missiles will add commonality to U.S. Air Force and Navy fighters, thus helping to reduce procurement and maintenance costs. The rail launcher uses common modules to allow an interchange of parts between launchers on different aircraft. Hughes is building more than 600 launchers for Air Force F-15 and F-16 aircraft. Eventually the launchers will be installed on the Navy's F-14 and F/A-18 aircraft. In addition, preliminary work is under way to adapt the launchers to Britain's Tornados and Sea Harriers.

The Australian Army will use a radar simulator to train operators and maintenance personnel on the AN/TPQ-36 Firefinder weapon locating radar. The trainer, designed by Hughes and built by British Aerospace Australia, is a computerized system that trains personnel without using either the production radar or live artillery fire. The radar itself pinpoints the position of enemy mortar, artillery, and rocket launchers. It rapidly scans the horizon with a pencil-thin beam, forming an electronic curtain across the battlefield. After detecting incoming projectiles, the system backplots their trajectories and passes the data to friendly forces for counterfire.

Lasers will help halve the cost of inspecting metal parts for fighter aircraft radars when a new manufacturing technique goes into effect at Hughes. Advanced optics technology will be used to inspect newly fabricated radar antenna plates in three dimensions. Besides lowering costs, the process will reduce errors. The project is part of an Industrial Modernization Incentive Program (IMIP) awarded by the U.S. Navy and Air Force to help create the electronics factory of the future. IMIP is a share-the-savings concept that will reduce costs of the F-14, F-15, and F/A-18 Hornet Strike Fighter radar programs by more than $10 million, while improving the quality and reliability of the sytems.

Hughes needs college graduates with degrees in EE, ME, physics, computer science, and electronics technology. To find out how to become involved in any one of the 1,500 high-technology projects, ranging from submicron microelectronics to advanced large-scale electronics systems, contact Corporate College Relations Office, Hughes Aircraft Company, Dept. S2, C1/C128, P.O. Box 45066, Los Angeles, CA 90045-0066. Equal opportunity employer. U.S. citizenship required.

For more information write to: P.O. Box 45068, Los Angeles, CA 90045-0068

HUGHES
AIRCRAFT COMPANY

Subsidiary of GM Hughes Electronics

ELDOR

The Driving Sp

The quality is readily apparent. The comfort and convenience are pure Cadillac. Yet the 1987 Eldorado delivers an added dimension, the exhilaration of commanding a luxury car that loves the road. As you settle comfortably behind the wheel, you note that everything is close at hand, ready for you to take control. Cadillac convenience features let you know you're driving a special personal coupe. A coupe backed by a new longer limited warranty.

This Cadillac takes your driving pleasure very seriously.

The power of a transverse-mounted, fuel-injected V8 with the traction of front-wheel drive. The controlled ride of four-wheel independent suspension with Electronic Level Control. The assurance of four-wheel disc brakes. And the response of power-assisted rack-and-pinion steering. All of these make Eldorado a true driver's luxury car.

A personal environment designed around you.

Adjust the leather-wrapped tilt-and-telescope steering wheel. Position the six-way power driver seat. Set the power rearview mirrors. This Cadillac is designed to accommodate your individuality while it encourages and rewards driver involvement.

*In some cases, a deductible applies. See your dealer for details.

RADO
it of Cadillac.

Technology that enhances your control.

Eldorado acknowledges that information translates into driver control and confidence. Within its available American walnut instrument panel, an advanced multi-function Driver Information Center monitors everything from engine rpm to the outside temperature.

The confidence you feel behind the wheel is also aided by Eldorado's sleek aerodynamic styling, which contributes to remarkable road stability.

Technology that enhances your convenience.

Of course, Eldorado has all the comfort and conveniences that make a Cadillac a Cadillac. But this Eldorado uses leading-edge technology to further enhance these Cadillac hallmarks. Witness the sun-load sensor that measures the amount of sunlight entering the passenger compartment and then automatically instructs the Electronic Climate Control to adjust itself.

Eldorado's new 5-year/50,000-mile limited warranty.

Another word for a company's confidence in its product's quality is warranty. And another word for quality is Cadillac. Eldorado is backed by a new longer 5-year/50,000-mile limited warranty.*

Your Cadillac dealer invites you to experience the joy of the open road in Eldorado, the driving spirit of Cadillac.

LET'S GET IT TOGETHER...BUCKLE UP. GM

Congress in 1984, recently listed the exploration of Mars as one of several space projects the U.S. might undertake. The Reagan Administration is now considering its response (which the 1984 law requires it to make) to the commission's report.

Yellow Rain: A Few More Drops

Canadian findings that came to light earlier this year concerning the alleged use of tricothecene toxins as a biological weapon in Southeast Asia have been said by Matthew Meselson of Harvard University and other investigators to deliver a serious blow to the credibility of such charges. In the September "Science and the Citizen" an item reporting these developments got the critics' conclusion right but the Canadian government's facts wrong.

Here, from Ottawa and for the record, are the facts. The Canadian government conducted two studies, one in 1982 and the other in 1984. In 1982 physicians from the Department of National Defence were collecting epidemiological data on the military use of tricothecenes. Learning of a possible attack, they gathered such samples from the area as leaves, scrapings taken from dwellings and physiological specimens taken from individuals. The material was then dispatched to the Defence Research Establishment Ottawa (DREO). Establishment officials had the samples screened at government laboratories for tricothecenes. "No amounts of tricothecenes," according to the DREO final report, "that were significantly greater than those reported world-wide on stored cereal crops were found in any of the naturally-occurring materials although the conditions of collection and transport of samples from Southeast Asia to Canada do not preclude there having been higher levels present at the time of sampling."

A plastic bag given by a villager to the Canadian investigators did show relatively high levels of tricothecenes. "However," the report says, "the authenticity of the source of the bag cannot be verified as it was not collected by the Canadian team."

All the physiological samples gathered by the investigators showed no tricothecenes down to levels of 100 parts per billion.

In 1984 another Canadian agency, the Department of External Affairs, collected 270 blood samples in Southeast Asia from areas that had not been exposed to attack. According to a department official, a few of the individuals from whom samples were taken said they had been exposed to "yellow rain." All the samples were analyzed by workers in a government laboratory. Five of the 270 blood samples showed small quantities of tricothecenes above a level of one part per billion; none of these samples was from a self-professed attack victim. The few individuals who reported that they had been in an attack "a long time before" showed no traces of the toxins. The laboratory also analyzed 10 control samples, some of which had been spiked with toxin.

Where does this set of facts leave the debate? That seems to depend on one's point of view. Canadian officials who have been asked for more information by both journalists and U.S. officials maintain steadfast neutrality: "We won't go beyond what we can say scientifically," a spokesman stated. Critics of the Reagan Administration's position consider their contention sustained. For its part the Administration refuses to come in out of the rain.

PHYSICAL SCIENCES

Jumping Jack Flash

Is quantum mechanics a precise description of the inherently probabilistic behavior of individual particles? Or is it merely a prediction of the statistical rules that will be followed by large ensembles of such particles? The question is hard to answer experimentally because it is extremely difficult to isolate individual particles or to observe single quantum "events." Three groups have now come as close as is possible to observing a single occurrence of one of the most fundamental of quantum events: the transition of an electron from one energy level to another within an atom.

The investigators were Warren Nagourney and his colleagues at the University of Washington, Th. Sauter and his colleagues at the Institute for Experimental Physics at the University of Hamburg and James C. Bergquist and his colleagues at the U.S. National Bureau of Standards. They all report their work in *Physical Review Letters.* The three groups took essentially the same approach; the procedure followed by the NBS group is a good example of the general technique.

The NBS workers leaked a small amount of mercury vapor into a vacuum chamber and ionized some of the mercury atoms (stripped an electron off each of them, thereby giving them an electric charge) with an energetic beam of electrons. They then caught the ions in an electromagnetic trap known as a Paul trap, which consists of a complex arrangement of oscillating electromagnetic fields. They cooled them by a method known as laser cooling, in which light is bounced off ions in order to slow them.

The center of the Paul trap was exposed to a wavelength of light that caused the ions to fluoresce: each ion repeatedly absorbed photons and re-emitted them in different directions, while its outermost electron jumped rapidly between the ground state (its state of lowest energy) and a much higher, excited state. A fluorescing atom scatters a beam of light and stands out in it much as a speck of dust scatters and stands out in a sunbeam. The investigators were able to determine how many ions they had caught by gauging the intensity of the stimulated fluorescence.

When they found they had caught a single ion, they continued to bathe it in the light that stimulated fluorescence, but they also added a weak beam of light at a wavelength that could stimulate the electron to jump to a state whose energy is between that of the ground state and the much higher state. When the electron is in the intermediate state, it cannot fluoresce. The mercury electron tends to remain in the intermediate state for a comparatively long time (about a tenth of a second, as opposed to the roughly two-trillionths of a second it remains in the higher state); during that time it is not fluorescing and is therefore dark.

Whenever the electron jumped into the state of intermediate energy, the ion stopped fluorescing. When the electron dropped back down again, the ion resumed fluorescing. The ion thus acted almost as an indicator light: it glowed whenever it was cycling between the ground state and the high state and remained dark when it was in the intermediate state.

The ability to detect single quantum jumps provides physicists with a tiny but powerful light with which to illuminate some of the most fundamental questions in physics.

How I Wonder What You Are

Like a well-hit clay pigeon, a recently discovered celestial object that was thought to be the first planetlike body ever sighted has vanished from the skies. Its disappearance raises doubt that the body ever existed.

In 1984 Donald W. McCarthy, Jr., and Frank J. Low of the University of Arizona and Ronald G. Probst of the National Optical Astronomy Observatories reported finding a planetlike "brown dwarf" circling the faint star Van Biesbroeck 8 (VB 8), some 21 light-years from the earth. Brown dwarfs, which had been postulated but not observed, are intermediate be-

tween stars and planets: they are cooler, dimmer and less massive than stars but hotter than planets. Nevertheless, because the newly discovered body, named Van Biesbroeck 8B (VB 8B), was not a star and appeared to be in a planetlike orbit around a star, McCarthy and his associates concluded that the object could be considered a planet (see "Science and the Citizen," April, 1985). To observe VB 8B in 1984, the Arizona group employed speckle interferometry, a technique that combines a series of short exposures into one image to overcome blurring caused by the earth's atmosphere.

In a paper scheduled for publication in *Astrophysics Journal Letters* Michael F. Skrutskie of Cornell University and William J. Forrest and Mark A. Shure of the University of Rochester report that they have not been able to find VB 8B in its expected location. French astronomers who were working at an observatory in Chile also have failed to find the planet.

Because speckle interferometry is a somewhat indirect means of visualizing celestial objects, Skrutskie, Forrest and Shure decided to observe VB 8B more directly, through a sensitive infrared array detector mounted on a three-meter telescope at the National Aeronautics and Space Administration's Infrared Telescope Facility on Mauna Kea in Hawaii. If VB 8B existed, the workers say, it should have shown up clearly in the images produced by the detector. It did not.

McCarthy agrees that VB 8B is not now in its expected location. Speckle interferometry, he says, may have produced misleading images. On the other hand, he notes, "no one can rule out the possibility" that VB 8B has moved very close to Van Biesbroeck 8 and hence can no longer be seen. Such movement could indicate that VB 8B is both bigger and moving faster than originally thought and is therefore actually a low-mass star.

Forrest offers a third explanation. He suggests that the Arizona group may actually have sighted a new kind of nebula, a cloud of gas or small particles. Forrest and his colleagues did see a diffuse image that might have been either a nebula or an artifact of their technique. Until recently it was thought that nebulas did not surround cool, low-mass stars such as VB 8.

Do the new findings mean there are no brown dwarfs, and no true planets outside the solar system? Forrest and his colleagues note that brown dwarfs have been returned to "the realm of theoretical speculation," but Forrest himself suspects that undiscovered planets are probably abundant. A planet outside the solar system, he adds, would be very difficult to detect from the earth because it would be small, dark and distant.

Red Rovers

Of the 10,000 or so meteorites that have been collected and analyzed, eight are particularly unusual. They are so unusual, in fact, that since 1979 some investigators have thought they might have originated not in asteroids, as most meteorites did, but on the surface of Mars. The problem with this hypothesis is that the meteorites have not been subjected to the enormous shocks that would seem to have been necessary to accelerate them to Mars's escape velocity, which is greater than five kilometers per second (about 11,000 miles per hour). John D. O'Keefe and Thomas J. Ahrens of the California Institute of Technology may have solved the problem by proposing a comparatively gentle mechanism that could have ejected pieces of the Martian surface into interplanetary space.

The meteorites are called SNC meteorites, after the regions where three of them were found: Shergotty in India, Nakhla in Egypt and Chassigny in France. SNC meteorites consist of rocks that crystallized about 1.3 billion years ago; that is much more recent than the times when asteroids are thought to have cooled (about 4.5 billion years ago). The meteorites' composition and texture indicate they formed on a planet that has a strong gravitational field. Many of them have concentrations of volatile elements consistent with analyses of the Martian soil, and glassy inclusions in the meteorites (which presumably formed in the extreme heat of whatever process ejected them from a parent body) have trapped the noble gases argon, krypton, xenon and nitrogen in the same relative abundances as the Viking missions found in the Martian atmosphere.

The hypothesis that SNC meteorites originated on Mars runs into trouble because even though a large body colliding directly with the planet's surface could have hurled chunks of material into space, the pieces would have been completely melted or vaporized by the shock and would therefore not resemble SNC meteorites, which are only very lightly shocked. How, then, could the SNC meteorites have separated from the Martian surface?

Writing in *Science,* O'Keefe and Ahrens describe how they devised a hypothesis that answers the question. With a high-speed gun, the investigators fired projectiles at samples of planetary crust and fed the resulting data into a series of computer models. They found that a large object (from .1 to one kilometer in radius) striking the Martian surface at oblique angles between 25 and 60 degrees would vaporize large amounts of the object itself and of the Martian crust and the water and carbon dioxide vapor trapped in it, forming a jet of high-velocity gas. The jet would entrain, or sweep up, pieces of the crust and blow them beyond the planet's gravitational field, perhaps melting their outer surfaces. The fragments would then go into orbit around the sun, and a few of them might cross the earth's path.

TECHNOLOGY

The Eyes Have It

"Nature provides her creatures with the best possible systems; the engineer has to try to implement them in devices," said Robert Zinter of the Institute of Optics at the University of Rochester. Zinter was describing the source of the concept for a distortion-free, wide-angle optical system he has begun to develop.

A typical wide-angle lens produces a distorted image because it gathers light along one optical axis from an extended scene and focuses it on a plane. Therefore the parts of the image that are off-axis, at the edge of the field of view, are distorted. The result is the classic "fish eye" effect.

To solve the problem Zinter turned to a natural model: the compound eye that has evolved in insects and some crustaceans. Such an eye consists of thousands of tiny rodlike lenses known as ommatidia. Each ommatidium accepts light from the environment and focuses it. Because the ommatidia are aligned so that there is a small angle between the axes of the lenses, the images they produce overlap one another. And because each ommatidium delivers a narrow field of view along its elongated optical axis, there is very little distortion in its image—and in the composite image derived from the entire array of ommatidia.

In place of ommatidia Zinter employs optical rods a millimeter in diameter, which he arranges in a hemispherical or fanlike pattern. The rods are treated by ion diffusion so that each rod behaves like an ommatidium and focuses light. Specifically, the index of refraction in a rod is highest at the center and decreases radially. As a result light from a wavefront impinging on the rod at the center has a lower velocity than light impinging at some distance from the rod's center; a wavefront moving forward along a straight line therefore becomes convergent and

focuses to produce an image. Like the compound eye, a bundle of such fibers impinging on a hemispherical receptor produces an image that is virtually free of distortion.

Zinter has constructed a two-dimensional prototype of his system that has yielded crude but undistorted images of a parking lot near the institute. A fully functional version, he says, could have applications in robotic manufacturing systems, surveillance devices and target-acquisition devices.

Mother of Necessity

Holography has seemed for years to epitomize an invention looking for a necessity. The technique, by which strikingly vivid three-dimensional images can be created, enables a viewer to inspect a representation from almost any angle as if it were the thing itself, standing independently in real space. To create a hologram an object (or assemblage of objects) is bathed in monochromatic light (generated by a laser); light reflected from the object mixes at the surface of a photographic emulsion with a reference beam. Reinforcement and cancellation between the wavefronts of the beams create an interference pattern on the film. When the transparency is illuminated, the process that created the interference pattern is reversed and a three-dimensional image appears. Although holography has found limited, hi-tech industrial applications, more widespread use has eluded it.

A group at the Media Laboratory of the Massachusetts Institute of Technology may be on the verge of changing all that. Under the leadership of Stephen A. Benton, the team has begun to adapt holography for use as a form of computer-information display. Largely supported by funding from the General Motors Corporation, Benton and his colleagues have programmed a computer to create holographic images of a Camaro; medical interest in the technique has also led them to image a pelvic bone.

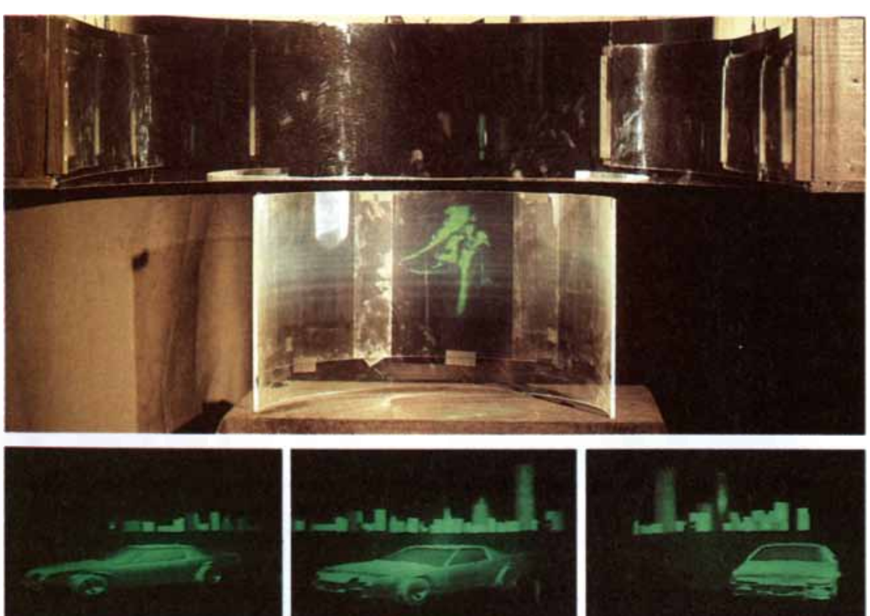
To create a hologram the computer stores a list of 10,000 points that define an object's entire surface in three dimensions. By means of projective geometry the computer then mathematically synthesizes 956 images, each of which represents a view from a somewhat different perspective. Each image, precorrected for distortions it will be subject to during processing, is displayed on a television monitor and then photographed on 35-millimeter cinematographic film.

After the film has been processed a laser beam projects each transparency in turn onto a ground-glass screen. Light from the image on the screen mixes with a reference beam at the surface of the holographic film. When the film is illuminated, the hologram (Camaro or pelvic bone) appears, floating before the viewer.

Benton and his colleagues are attempting to sharpen the images they produce and increase the speed with which the system operates. Ultimately they hope to replace the cinematographic film with a liquid-crystal television display from which the hologram would be made directly.

Such a system would have a wide range of applications, according to Michael A. Teitel, a member of the group. It would help to reduce sharply the cost and time required to design an automobile by eliminating the need to make clay or wood replicas. Physicians, by feeding CAT-scan data or other visual information into the system, could accurately reconstruct a lesion affecting an organ or a bone. Architects and city planners could use the technique to model a structure and gauge its impact on the appearance of the surrounding cityscape.

Perhaps someday PAC-men may even gobble in 3-D.



COMPUTER-GENERATED HOLOGRAM at M.I.T.'s Media Laboratory displays a failed artificial hip joint (*top*). The image is generated by illuminating a holographic film (on the curved plastic screen) from behind with a laser, whose beam is directed by the curved mirror above the screen. The other images show three views of a car (*bottom*).

Superglass

Investigators at the University of Rochester have found a way to make the glass employed in many commercial lasers stronger and more resistant to cracks than the glass now in use. Consequently lasers can be fired at higher repetition rates and run at a significantly (up to six times) greater power than is now possible. Because existing glass, which serves as a matrix for neodymium and other compounds that lase, can easily be replaced with the new glass, it should be possible to quickly improve the quality of existing lasers having applications in such diverse fields as industrial machining, ocular surgery and microchip fabrication.

In developing the strengthened glass the Rochester workers, led by Kathleen A. Cerqua and Stephen D. Jacobs, collaborated closely with Kigre, Inc., of Toledo, Ohio, a company that manufactures slabs of laser glass. The slabs, measuring 160 by 15 by eight millimeters, contain a large amount of the element lithium.

The collaborators immersed samples of the glass in a bath of molten salt containing atoms of sodium and potassium. Even though the two elements are much larger in size than lithium, they have similar chemical properties. As a result, when a slab is left to soak in the bath for several days, atoms of lithium diffuse from the surface of the glass and are replaced by sodium and potassium. When they are squeezed into the small "holes" left by lithium, the larger atoms create a compressive layer of stress around the slab. The layer is typically only 60 micrometers (thousandths of a millimeter) thick, but the compression it generates is sufficient to hold the slab together and to prevent cracks from

AT&T BRINGS YOU A PROGRAM THAT CONTINUES TO MAKE HISTORY.

Three years ago, The MacNeil/Lehrer Report became The MacNeil/Lehrer NewsHour—and made history as the first hour-long evening news program on national television.

Through in-depth analysis and interviews with expert guests, Robert MacNeil and Jim Lehrer have been helping many of us better understand the complex world we live in.

We at AT&T began supporting the MacNeil/Lehrer approach to the news in 1979. We're still doing it today.

We're proud of the association. We believe it's one more reflection of our commitment to excellence. In business. And in the quality of life.

The MacNeil/Lehrer NewsHour. Weeknights on PBS. Have a look for yourself. See why we're so proud.

AT&T
The right choice.

traveling through it. Initial tests of the strengthened glass have been successful, and the development and marketing of a commercial product are now being explored.

BIOLOGICAL SCIENCES

Tags for Destruction

Why do some proteins last as long as the cells that made them, whereas others are broken down in minutes? That the variation in protein lifetime is crucial to cell function is clear. Most structural proteins, for example, must be long-lived, but regulatory proteins undergo rapid swings in concentration, which implies quick degradation. Just how the cell singles out certain proteins for destruction while sparing others has long been an enigma. Andreas Bachmair, Daniel Finley and Alexander Varshavsky of the Massachusetts Institute of Technology may have happened on a major part of the answer: a simple code specifying protein lifetime.

Writing in *Science,* the M.I.T. workers tell how they focused at first on ubiquitin, a mysterious protein that, as its name implies, is ubiquitous in the cells of higher organisms. Earlier studies by Varshavsky and others had indicated that ubiquitin must be coupled to proteins if they are to be degraded. To test the possibility that ubiquitin acts to mark proteins for destruction, the group combined the gene for ubiquitin with a gene for a bacterial enzyme, which served as a test protein, and introduced the resulting chimeric gene into yeast cells. The workers expected the gene would be expressed as a ubiquitin-containing "fusion protein"; its lifetime could then be compared with that of the ubiquitin-free test protein. One of the yeast's own enzymes thwarted the plan by quickly snipping the ubiquitin from the test protein.

In an effort to foil the enzyme by disguising the junction between ubiquitin and the test protein, the workers altered the part of the chimeric gene coding for a single amino acid (one of the chemical units of proteins) at the start of the test protein, next to the junction. The strategy failed; ubiquitin was snipped away regardless of the initial amino acid.

The failed strategy nonetheless had a felicitous consequence. Newly synthesized proteins always begin with the amino acid methionine, because the code for methionine is also the signal for a cell's genetic machinery to start protein synthesis. In living systems later processing exposes other amino acids, but investigators had never been able to harness the mechanisms in the laboratory and vary the initial amino acid at will. Now, by in effect using ubiquitin as an expendable starting sequence, the M.I.T. group was able to do just that.

The fortuitous new ability unexpectedly revealed a code for protein degradation, one in which ubiquitin seems to have no part. In attempting to deceive the yeast enzyme, the workers produced versions of the test protein that had 15 of the 20 amino acids at their amino terminus, or starting position. These otherwise identical proteins varied enormously in their stability within the yeast cells; their half-lives ranged from less than three minutes to more than 20 hours. Each amino-terminal amino acid was associated with a specific half-life, a correlation the authors refer to as the "N end" (for amino-terminus) rule.

Does the N-end rule hold for other proteins as well? An examination of the proteins whose stabilities in living cells and amino termini are known suggested it does. Stable proteins invariably begin with one of the amino acids that confer long life under the N-end rule. The one short-lived intracellular protein for which the amino terminus is known also follows the N-end rule: it has a half-life of three minutes, and it begins with the amino acid that is associated with the shortest half-life.

Executing this remarkably simple code for protein degradation calls for complex cellular machinery, in which ubiquitin is expected to have a role after all. An unidentified "N-end-reading" enzyme would be needed to examine amino termini and select proteins for destruction according to the N-end rule. Other enzymes would then degrade them. Varshavsky thinks that even though ubiquitin plays no part in the code itself, it must be coupled to a protein that has already been selected for degradation in order for the actual breakdown to take place.

Tumor in Recession

By now the concept of an oncogene—a gene whose malfunction causes cancer—has become familiar. By and large, however, the oncogenes reported so far have been dominant: a change in the gene activates a normally silent stretch of DNA whose activity begins the chain of events leading to a tumor. Yet for some time it has been hypothesized that there is another type of oncogene, a recessive oncogene. Its cancer-causing effects would result from its being inactivated rather than being abnormally activated. Stephen H. Friend of the Whitehead Institute for Biomedical Research and his colleagues have now isolated the first recessive oncogene, which apparently causes tumors in eyes and in bone.

The specific tumors Friend and his co-workers investigated are called retinoblastomas and osteosarcomas. Both are rare. The retinoblastomas generally appear within the first year of life and are treatable; the osteosarcomas usually appear at puberty and are more aggressive. Thaddeus P. Dryja of the Harvard Medical School had already identified the approximate location of the defect that gives rise to these tumors. It is a small DNA segment in the band called q14, which is found on the long arm of human chromosome 13.

Friend, collaborating with Dryja, Robert A. Weinberg of the Whitehead Institute (who has done much of the work on dominant oncogenes) and others, began assembling a set of DNA fragments spanning the region Dryja had identified. When the set of ordered fragments was matched with the DNA of tumor cells, it was found that the tumor cells lack variable pieces of DNA in a section of q14 that now appears to be the retinoblastoma (*Rb*) oncogene. What is more, the tumor cells lack the RNA message corresponding to the *Rb* gene, which suggests the deletions have indeed inactivated the gene.

How might a recessive cancer gene work? "This is, of course, all very speculative," Friend said, "but if there is a gene that ordinarily tells the cell to stop dividing during the process of development, and if that gene were inactivated, then instead of coming to the end of the genetic program and differentiating into a specialized cell type, as normally happens, the cell might be left to make a clone of undifferentiated cells." Such a clone (a group of genetically identical cells) is thought to be the precursor of many types of tumor. "The normal function of the gene," Friend continued, "would thus be to tell the cells when to stop dividing."

And how common will recessive oncogenes turn out to be? Friend, whose results were published in a recent issue of *Nature,* was cautious. "I can't say that this has any relevance to major killers such as lung cancer or breast cancer," he said, "but its relevance certainly extends beyond the two tumors we worked with. There is a group of tumors that are known to have a familial predisposition—a pattern of inheritance similar to that of retinoblastoma or osteosarcoma—and to involve chromosomal deletions. Among them are certain subtypes of leukemia and some solid tumors of children. I think that, at a minimum, recessive oncogenes will turn out to be significant in this group."

MEDICINE

Heart Attack

Medical research has entered the era of the large-scale, high-technology project. In all areas of research such projects are attended by certain crucial questions: Who should decide whether or not a program is to be funded? How should its progress be measured and the benefits evaluated? In medicine ethical questions come up as well: What kinds of experimentation on human beings should be done and when? Are there programs that would be more cost-effective in improving the health of the nation as a whole? To what degree should private companies be allowed proprietary benefit from projects initiated and sustained by Federal funding? Writing in *Medical Heritage,* Barton J. Bernstein, a historian at Stanford University, argues that the history of the program to develop an artificial heart illustrates how such questions should not be answered.

The Government's decisions to fund and continue the project, Bernstein writes, have been based on highly overoptimistic recommendations submitted by committees made up largely of biomedical "insiders" rather than on the basis of open public dialogue. The committees assumed that technological solutions could be applied to extremely difficult problems, many of which remain unsolved more than 20 years later, and relied on "rough guesses" to estimate the number of people who could be helped by an artificial heart and to project its final cost.

Bernstein states the committees paid limited attention to issues that cannot be quantified, such as the legal and ethical implications of artificial hearts and the quality of life recipients could anticipate: apart from physical inconvenience, there was the likelihood that psychological readjustment to family and work would be hard. (Bernstein notes that the incidence of suicide among dialysis patients is roughly seven times the normal rate.)

Bernstein also criticizes the program on the basis of its cost. He points out that the care and treatment of Barney Clark, the first patient to receive an artificial heart with the approval of the Food and Drug Administration, cost more than $250,000, and that it would have been more expensive if doctors had not waived many of their fees. In 1985 an advisory panel of the National Heart, Lung, and Blood Institute estimated that the average cost to each artificial-heart recipient will eventually be about $150,000. Who will pay these costs? What will be the effect on the medical economy? Bernstein also suggests that dollar for dollar more benefit might have been gained by a large-scale promotion of preventive medicine (such as an antismoking campaign) than was gained by the large amounts of Federal money spent on artificial-heart research.

Bernstein's exploration of history has raised some contemporary hackles. When asked about the cost of the artificial heart, William C. DeVries, the surgeon who implanted Clark's heart and who has since been hired by the Humana Heart Institute to continue implant operations, replies: "Right now it is totally inappropriate to evaluate what the cost will eventually be.... This is phase-one research. I'm not trying to cut costs now. Now I do the most expensive thing, to learn what works in handling the patient.... It is inappropriate to limit science because of cost."

Another of Bernstein's criticisms is that DeVries and others are implanting hearts in human beings even though there have been serious medical problems in the calves on which the heart has been tested. Bernstein asks: "Was [the Clark case] basically an experiment? Was there really any reasonable expectation that Barney Clark was going to live a comfortable life?" Citing the large number of strokes and other complications in artificial-heart recipients, Bernstein argues that implanting the artificial heart in patients at such an early stage raises troubling issues of experimentation in human subjects. DeVries responds that some questions cannot be answered in animal models: "Just because you have complications in animals doesn't mean you stop clinical use. You need to ask: Have you anything more to learn from animals?"

Perhaps the most disturbing issue Bernstein addresses is that of profits and secrecy. In August, 1984, DeVries moved from the University of Utah to Humana, a for-profit institution; the Jarvik-7 heart he implants is made by a private firm, Symbion, Inc., Bernstein notes, but it was developed with considerable Federal funding. Outside requests to review DeVries' research protocol have been denied, as have requests for data from implanted patients, which might help outsiders to evaluate the function of the device. "Such troubling secrecy, especially when added to the shift of the project to a for-profit hospital," writes Bernstein, "raised additional questions among critics about scientific objectivity, independent scrutiny, the selection of patients, and manipulation of the press.... Symbion and other companies, even though they have benefited handsomely from Federal funding for this research, want to keep much of what they have learned secret."

Richard W. Alder, executive vice-president and chief operating officer of Symbion, denies charges of secrecy. "The protocols are available to the Food and Drug Administration and to all principal investigators. We limit the distribution because it is highly technical and constantly changing, and we don't see the purpose of publishing.... It's confusing to release a lot of data when it's all evolving. It is released from time to time by principal investigators." DeVries says that he is preparing several journal articles describing various protocols, but that in general the protocols are not released to the public.

Surrounding all these specific criticisms Bernstein sees a larger issue. He reflects on national technological decision making as a whole and asks: "Why is it that technological 'fixes' or 'solutions' seem more attractive than preventive activities?" He suggests it is because prevention does not involve a dramatic event or a hero. He also proposes that decisions to commit large amounts of funding to technological projects should be more broadly based and should rely on those outside the immediate field of concern.

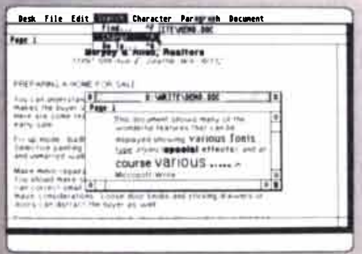
RU-486

The National Institute of Child Health and Human Development has a hot new drug to deal with. The agent, RU-486, appears to be effective against Cushing's syndrome and may retard breast cancer; it shows promise for treating glaucoma, arresting lymphomas and improving fertility by inhibiting endometriosis. Paradoxically, RU-486 may be an effective once-a-month contraceptive, an attribute that also makes the drug a sure-fire igniter of controversy. Right-to-life groups, which regard RU-486 as an abortifacient, have been pressing the Administration to halt studies of the drug and keep it off the market.

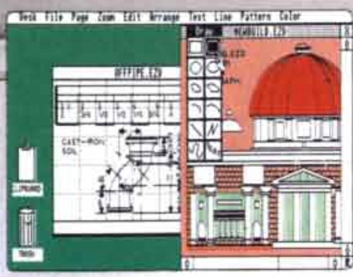
RU-486 was developed by Roussel-UCLAF, a French pharmaceutical company. Officials of the company's U.S. affiliate, the Roussel Corporation, are reluctant to say much about the drug but indicate there is no intention to seek approval for its distribution in this country.

That might be unfortunate, considering the drug's multifaceted potential. Its varied effects stem from its ability to interfere with important

T H E **S T** C O M P U T E R S



Word Processing



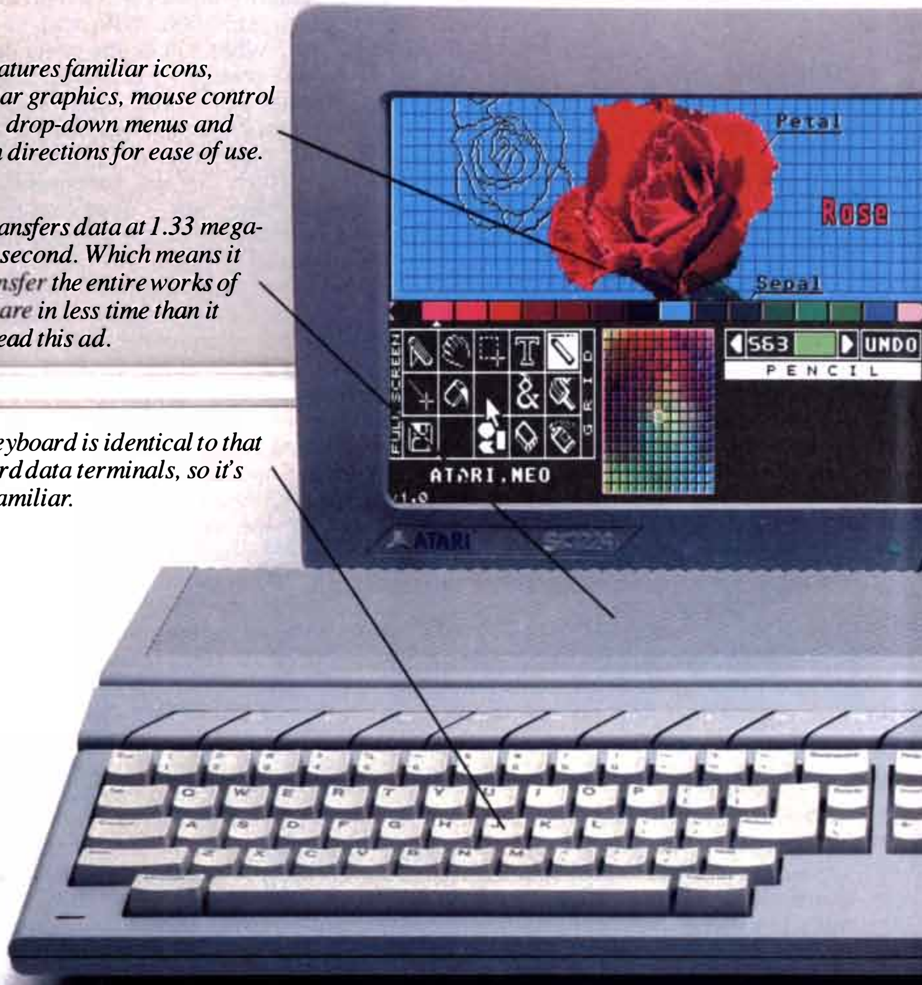
Graphics & Design

Music Composition

The ST features familiar icons, spectacular graphics, mouse control windows, drop-down menus and on-screen directions for ease of use.

The ST transfers data at 1.33 megabytes per second. Which means it could transfer the entire works of Shakespeare in less time than it takes to read this ad.

The ST keyboard is identical to that of standard data terminals, so it's already familiar.



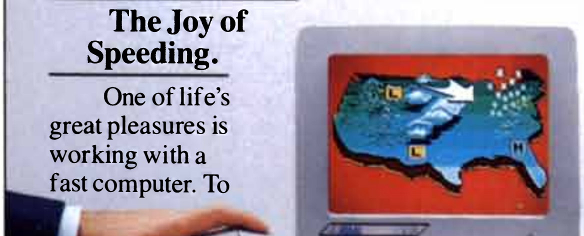
Introducing technology

Finally, there's a personal computer that not only solves problems like other computers, but also solves the one problem other computers created. Affordability.

Introducing the ST™ Computers from Atari: The 520ST™ with a 512K memory and the 1040ST™ with a full megabyte. The ST was designed utilizing the most recent breakthroughs in semiconductor technology, producing a PC that does more tasks with fewer parts.

Which means it costs less to make. And less to buy.

The Joy of Speeding.



One of life's great pleasures is working with a fast computer. To bring the ST up to speed, Atari starts with the Motorola 68000 chip—the same "brain" you'll find in the Macintosh.™ Then, Atari adds the extra oomph of four exclusive chips—specially designed to handle several functions simultaneously. (Other PCs limp along handling one function at a time).

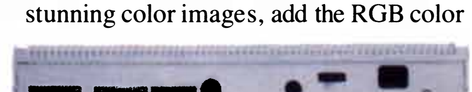
This results in making the ST much faster in the computing process. Faster in moving data within the system. Faster in getting information to the screen.

So now, you can run programs like word processing, database management, and financial planning with more zip and efficiency than ever before. A nice feeling.

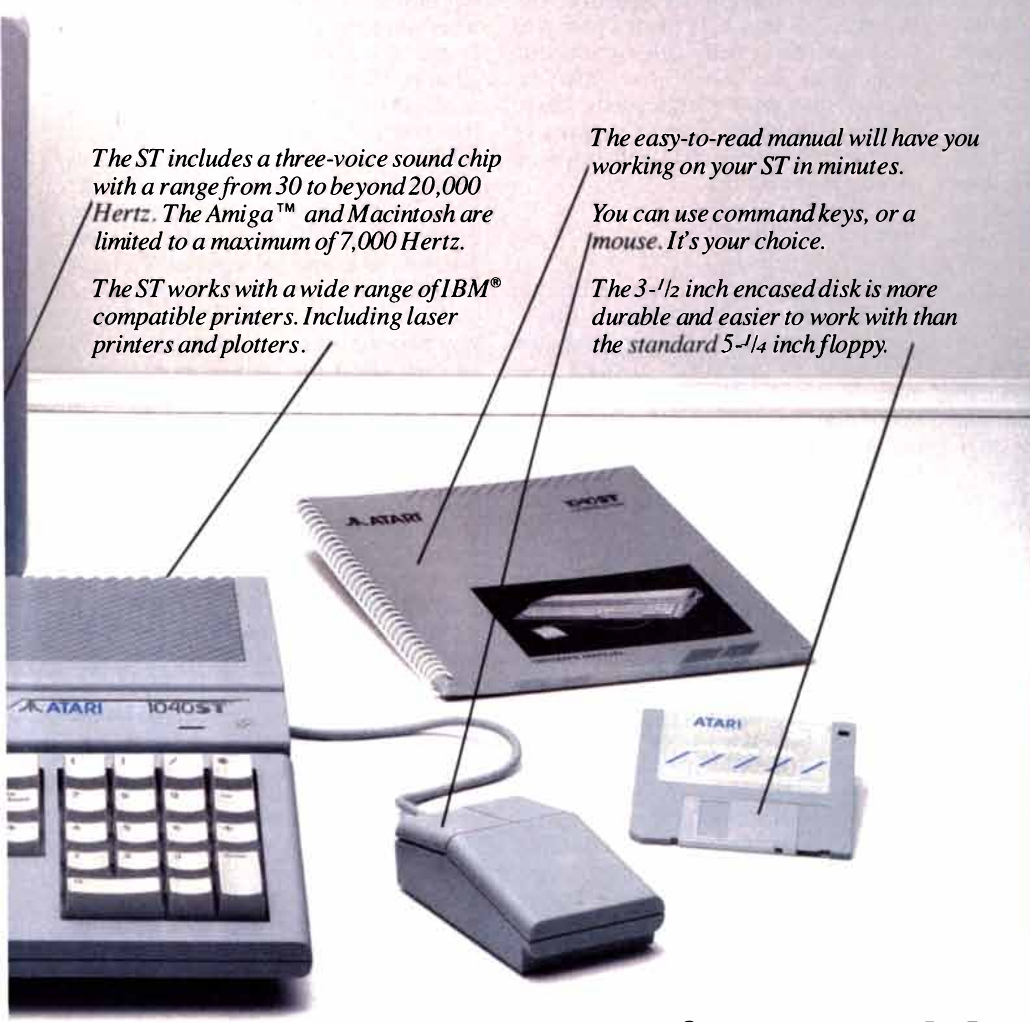
Compare Our Components.

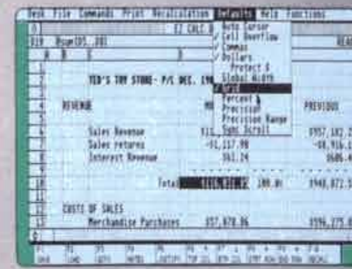
A computer is only the sum of its components. So we made each one better. Look at the layout of the ST keyboard, for example. You get a full numeric keypad. Plus a cursor control keypad with editing keys. Plus 10 programmable function keys. Now add the mouse and consider the options.

The monochrome monitor is a beauty. Taking its broad bandwidth signal from the ST's exclusive video chip, it displays a resolution of 640 × 400 pixels. This gives you razor-sharp, jitter-free text display for word processing and CAD work (very easy on the eyes). Or, for stunning color images, add the RGB color

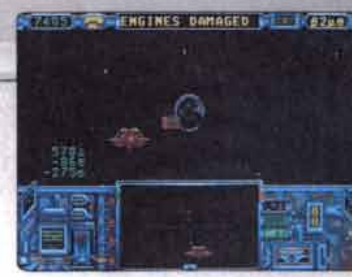


Atari, SH204, ST, 520ST and 1040ST are trademarks or registered trademarks of Atari Corporation. Activision is a registered trademark of Activision, Inc. Microsoft is a registered trademark of Microsoft Corporation. Macintosh is a trademark licensed to Apple Computer, Inc. Spinnaker is a registered trademark of Spinnaker Software Corporation.

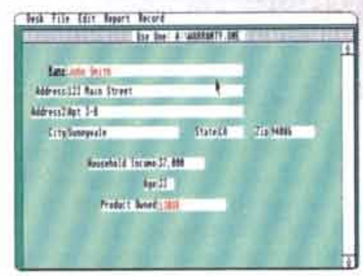




Spreadsheets



Entertainment



Database Management

so advanced, it's affordable.

nonitor. No interface ooard needed. Just olug it in. (Try doing hat with a Mac!)

You get an external disk drive with the 520ST. An internal, double-sided disk drive with the 1040ST. Both have a disk speed many times faster than previous PCs. And they're blissfully quiet.

Plus, many of the costly peripherals you have to add on with other PCs are already built into an ST. Like the built-in MIDI (musical synthesizer interface) port. And the industry-standard printer port and modem port. And for even more memory, a port for he SH204™ hard disk drive, with twenty megabytes of storage and the fastest transfer rate in the industry.

With hundreds of software programs already available, an ST can grow with your imagination, or your business. Companies like Microsoft,® Spinnaker,® Activision® and more are continually making contributions to the ST software library. And some popular programs originally designed for other computers are actually being upgraded to take full advantage of the ST's capabilities!

The Price of Power.

Best of all, the cost of an ST is so low, it may come as something of a shock. The 520ST sells for under $800, including monochrome monitor.* The 1040ST, with a full megabyte of memory, for under $1,000. That's less than one dollar a byte.

So now, you don't have to be rich to be powerful.

To see why Infoworld called the ST "The best hardware value of the year" check it out at your Atari dealer. For the one nearest you, call 1 800 443 8020. 9AM-5PM Mon.-Fri., Pacific Time.

*RGB color monitor, $200 additional.

THE ST™ COMPUTERS from ATARI®

chemical messengers. RU-486 is first of all an antiglucocorticoid agent: it blocks the receptors for such steroid hormones as cortisol. Investigators at the NICHHD report dramatic results after testing the drug to treat certain forms of Cushing's syndrome, a life-threatening disease that results from excess activity of adrenocortical hormones. The glucocorticoid-blocking action suggests the drug may provide an effective therapy for glaucoma induced by those hormones and for glucocorticoid-dependent lymphoma.

RU-486 also blocks the receptors for progesterone, the major steroid female sex hormone. In doing so it can prevent implantation: the attachment of a fertilized ovum, in an early stage of division, to the lining of the uterus, where it develops into a fetus. The institute is therefore interested in the drug's potential as a contraceptive agent that could be administered once a month to ensure menstruation and thereby prevent pregnancy.

The agent's antiprogesterone activity may provide a therapy for endometriosis (in which uterine tissue grows aberrantly on other reproductive organs) and thereby improve fertility in certain patients. In addition the drug is being considered as a weapon against breast cancer, the progress of which can sometimes be slowed by reducing the progesterone level.

In France and Scandinavia in particular, RU-486 is being investigated as a possible first-trimester abortifacient. In answer to inquiries from a group of U.S. senators allied with right-to-life proponents, the NICHHD has stated explicitly that no Federal funds "have been or are being used to support or conduct research on RU-486 as a drug to induce abortion."

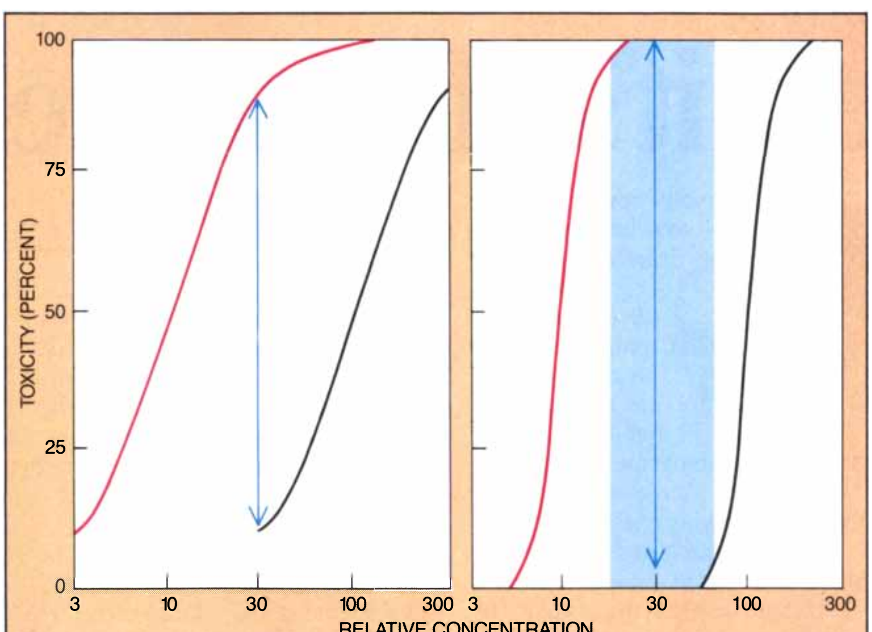
Binary Weapon

Nice discrimination between target and surround is a primary objective in cancer chemotherapy, in the treatment of parasitic diseases and even in weed killing. One wants an agent that attacks tumor cells but not normal ones, the parasite but not the host, weeds but not the crop or the farmer. What if two rather harmless substances could be administered that combine—primarily in the tumor cell, in the parasite or in the weed—to form a lethal compound? A very preliminary step toward achieving just such molecular teamwork has been reported in *Science* by Darryl Rideout of the Research Institute of Scripps Clinic, who works in a joint research program sponsored by the clinic and PPG Industries, Inc.

Rideout knew that some substances bind to tumor cells somewhat more readily than they do to other cells. Perhaps the selectivity could be amplified. One way would be to capitalize on the fact that if two or more different molecules must interact to achieve a certain biological response, the dose-response relation becomes steeper. That is, the effect on a tissue increases more sharply, for a given increase in the concentration of the substances, than it would for the same increase in the concentration of a single molecule. If a steeper dose-response rate can be attained, it should be translatable into a larger "therapeutic window": a larger difference, over a wider range of concentrations, between the effect a given concentration has on tumor cells and the effect the same concentration has on normal cells (*see illustration*). A self-assembling cytotoxin, or cell-killing substance, formed by the combination of two less toxic molecules should exhibit the required synergism.

To test the concept Rideout first showed that two mildly cytotoxic chemicals called decanal and AOG would react with each other to form a third substance. Then he tested them for synergistic cytotoxicity. Neither substance alone had any significant effect on human erythrocytes (red blood cells). When erythrocytes were exposed to a mixture of the two at concentrations of 28 micromoles per liter, however, they lysed (that is, their cell membrane dissolved and the cells burst) within 80 minutes. The rate of lysis increased with the concentration of either decanal or AOG, indicating that the two had indeed combined in situ to form a more toxic substance; that compound, called DIOG, was detected in solutions of lysed red cells.

When DIOG was applied directly to red cells at a concentration of only 14 micromoles per liter, it lysed them in 20 minutes, strongly suggesting that it was the effective cytotoxin. In a given length of time, however, the curve for the increase in lysis with increased concentration was much sharper for the combination of decanal and AOG than it was for their product, DIOG. In other words, a steeper dose-response curve was attained when the cytotoxin was introduced in the form of two components that self-assemble in situ. There should (at least in theory), then, be a large difference between the effect of a given concentration of pairs of self-assembling molecules on tumor tissue and on other tissue.



TWO HYPOTHETICAL DRUGS, both having tenfold selectivity for tumor cells, are compared. In the case of a drug that has a shallow dose-response curve (*left*), the difference (*blue arrow*) in cytotoxicity for tumor cells (*red*) and for normal cells (*black*) never exceeds about 80 percent: a concentration able to kill most tumor cells would harm many normal cells. In the case of a drug displaying a steep dose-response curve (*right*), the cytotoxicity difference is more than 95 percent over a sizable "therapeutic window" (*light blue*), within which a tumor-killing concentration should not harm normal cells.

THE NOBEL PRIZES

Physiology or Medicine

What controls the growth of an organism from a single fertilized egg into a complex system of billions of specialized cells? Part of the answer

lies in a class of signal proteins, called growth factors, whose first members were discovered in the 1950's by Rita Levi-Montalcini, now at the Institute of Cell Biology in Rome, and Stanley Cohen, now at the Vanderbilt University School of Medicine. For their collaborative efforts the two investigators shared this year's Nobel prize in physiology or medicine.

Levi-Montalcini discovered the first growth factor while working at Washington University in St. Louis. At the time she was studying the development of the nervous system in chick embryos. It had already been shown that a mouse tumor grafted onto a chick embryo would stimulate extensive nerve growth in the embryo. Levi-Montalcini found this to be true even when the tumor was not in direct contact with the embryo but was connected to it only by blood vessels; apparently, she later wrote, "the tumor was releasing some chemical factor that was in turn inducing the remarkable growth" of nerves in the embryo [see "The Nerve-Growth Factor," by Rita Levi-Montalcini and Pietro Calissano; SCIENTIFIC AMERICAN, June, 1979].

In 1953 Cohen joined Levi-Montalcini at Washington University, and it was he who succeeded in purifying the nerve-growth factor (NGF) and in showing that it was a protein. NGF, it is now known, is necessary for the survival during development of sympathetic nerve cells. It is synthesized in a wide range of tissues, and it appears to guide nerve fibers to the organs they serve: the fibers grow toward sources of NGF.

While studying NGF Cohen discovered a second growth factor, which he called epidermal growth factor (EGF). (It later emerged that EGF stimulates the growth not only of skin but also of many other cell types.) Cohen purified the protein and determined its amino acid sequence, and he also identified the receptor that binds EGF on cell surfaces. "For the first time," the Karolinska Institute noted in announcing Cohen's prize, "scientists had a factor available [that]...allowed studies of the growth process." Understanding normal growth, moreover, is the key to understanding pathological subversions of the process, including the uncontrolled proliferation of cancer cells.

Physics

Quantum mechanics, which is sometimes taken to be the embodiment of scientific arcana, lies at the heart of two eminently practical achievements that were recognized with the 1986 Nobel prize in physics: the electron microscope and the scanning tunneling microscope.

The award to Ernst Ruska, now retired from active research, was belated: Ruska built the first electron microscope while he was at the Technical University of Berlin in 1933. In so doing he was applying the quantum-mechanical postulate that an electron is both wave and particle. The wavelength of an electron is thousands of times less than the wavelength of visible light. Ruska and other workers knew the resolving power of an electron microscope would therefore be much greater than that of a light microscope—provided a way could be found to focus the electrons and produce an image.

Ruska's accomplishment was to develop magnetic coils that could act as electron lenses. Later, working for Siemens AG, he participated in the development of the first mass-produced electron microscope. Today the descendants of his device are ubiquitous, notably in biological and medical laboratories.

The scanning tunneling microscope, invented only a few years ago by Gerd Binnig and Heinrich Rohrer of the IBM Zurich Research Laboratory, exploits the wave nature of the electron in a different way [see "The Scanning Tunneling Microscope," by Gerd Binnig and Heinrich Rohrer; SCIENTIFIC AMERICAN, August, 1985]. Since electrons behave like waves, their positions are indeterminate; some of them actually "tunnel" out of an object and form a cloud above its surface. In the scanning tunneling microscope a fine stylus (its tip may consist of a single atom) is pushed to within a nanometer of a sample, so that the two electron clouds just touch. The application of a voltage causes an electric current, the tunneling current, to flow between the stylus and the sample.

Because the density of the electron clouds falls off exponentially with distance, the intensity of the tunneling current changes dramatically if the distance between the stylus and the sample is changed only slightly. The tunneling current can therefore serve to control a feedback mechanism that keeps the stylus at a constant height. As the stylus passes over the surface of the sample its vertical movements are measured. The vertical resolution of the measurements is about a hundredth of a nanometer, which means they can readily detect bumps in the surface made by individual atoms.

To function properly the scanning tunneling microscope must be impervious to the smallest external vibrations, and it is for solving such design problems that Binnig and Rohrer received the Nobel prize. Their device has immediate applications in the microelectronics industry. Beyond that, according to the Nobel committee, the invention has opened up "entirely new fields...for the study of the structure of matter."

Chemistry

If you heat two chemicals in an Erlenmeyer flask, you may produce a third. You will then know that substance *A* has reacted with substance *B* to form substance *C,* but you will not know what has actually happened on the molecular level: what energies the molecules of *A* and *B* had when they collided, for example, and whether *C* emerged directly from the collision or instead was formed from some unstable intermediate complex. For their work in providing "a much more detailed understanding of how chemical reactions take place," the Nobel prize in chemistry was awarded to Dudley R. Herschbach of Harvard University, Yuan T. Lee of the University of California at Berkeley and John C. Polanyi of the University of Toronto.

Herschbach pioneered the method of crossed molecular beams, in which two beams of molecules are accelerated to known energies and made to collide at a known angle. By measuring the energies and the angular distribution of the reaction products one can infer how pairs of molecules in the beams have interacted. Through this approach Herschbach has been able to analyze important types of direct reactions; he has also discovered indirect reactions that involve long-lived intermediate complexes. Lee, who initially worked in Herschbach's laboratory, refined the crossed-beam apparatus, making it possible to study complex reactions important in combustion chemistry and atmospheric chemistry.

Polanyi independently developed a new technique for analyzing the reaction products in crossed-beam experiments. In some chemical reactions excess energy is stored internally in the product molecules, which eventually emit the energy as infrared radiation. By measuring and analyzing the extremely weak infrared emissions, Polanyi has been able to infer how much of the total energy in a reaction is stored as vibrational energy and how much of it is imparted to the product molecules as kinetic energy of motion. The partition of the energy in turn yields information on the interatomic forces that drive the reaction. Polanyi's discovery of infrared "chemiluminescence" by vibrationally excited molecules also led to the development of chemical lasers.

M I T S U B I S H I

The sensation between here and there.

To satisfy those given to spontaneous fun often requires the most disciplined patience. For proof, consider the Mitsubishi Montero.

Montero's designed to give new freedom to your imagination. Patiently crafted with all the power of 2.6 liters, the independence of both 2- or 4-wheel drive and plenty of room for all your cargo. Even if all you're carrying is an impulsive imagination.

Montero's five-speed manual overdrive transmission and fully independent front torsion bar suspension help you navigate the shortcuts with confidence and ease. While its hefty all-terrain radials give you the ground clearance you need when you take the high road.

Of course in town, Montero still knows how to be a perfect gentleman. Its ergonomically shaped interior includes reclining front bucket seats and plush carpeting. And to make sure that cruising downtown is just as easy as cruising the backroads Montero comes equipped with power assist steering and power assist vented front disc brakes.

The Mitsubishi Montero. With prices starting as low as $9739,* it's not only a unique experience but an exceptional value. Patiently crafted to transport your imagination from here to there... even if there is nowhere.

*Mfrs. Suggested Retail Price. Taxes, license, freight, options, other dealer charges extra. Call 1-800-447-4700 for your nearest dealer.

Mitsubishi. Perfecting the Experience.™

MITSUBISHI MOTORS

M O N T E R O

The First Human Retrovirus

Part I of a two-part article on the human retroviruses. The first example, found in 1978, causes a rare leukemia. Its discovery laid the groundwork for identifying the related virus that causes AIDS

by Robert C. Gallo

Generally in nature the flow of genetic information is from DNA, where the information resides, to RNA, which serves as an intermediate, to proteins, which are the cell's functional molecules. Indeed, for many years it was believed that genetic information can flow only in one direction, and this proposition was known as the "central dogma" of molecular biology. Some time ago, however, the central dogma was subverted by the discovery of a group of biological objects that reverse the usual order of things. Called retroviruses, their genetic complement consists of RNA. They also contain an enzyme—reverse transcriptase—that uses the viral RNA as a template for making DNA, which integrates itself into the chromosomes of the host cell and there serves as the basis for viral replication.

The discovery of reverse transcriptase was exciting not only because it shed light on fundamental biological processes but also because it provided a handle for understanding how retroviruses cause disease. It was already known that these viruses can cause cancer in animals, mainly leukemias, which are tumors originating in white blood cells. Hence it seemed only logical to search for similar cancer-causing viruses in human beings. Yet the obstacles, both intellectual and technical, were so formidable that it was not until 1978, almost a decade after the discovery of reverse transcriptase, that my colleagues and I isolated the first human retrovirus.

We named our find human *T*-cell lymphotropic virus because it has an attraction for *T* lymphocytes, white blood cells with a crucial role in modulating the immune response. The virus does cause cancer: by entering a *T* cell, it can set in motion the chain of events leading to leukemia. (Indeed, for that reason it is sometimes called human *T*-cell leukemia virus.) Employing the techniques of epidemiology and molecular biology, we quickly found out a good deal about this pathogen and its mode of operation. What is more, we began to find its relatives. In 1982 we discovered a second human retrovirus, which also causes a leukemia. Instantly, human *T*-cell lymphotropic virus became HTLV-I, and the newcomer HTLV-II.

Even more surprising was the finding, in 1983 and 1984, that the agent of acquired immune deficiency syndrome (AIDS) is a related virus: HTLV-III. That discovery was surprising in part because AIDS and cancer have opposing effects. Whereas cancer entails uncontrolled proliferation of *T* cells, AIDS leads to their death, crippling the immune system. Thus within five years of the isolation of the first human retrovirus two distinct categories of such viruses had been found, each of which has quite different pathogenic effects. How these two categories came to be established is one of the most exciting stories of 20th-century biology. It will be told in two parts. The first part—this article—describes the discovery of HTLV-I. The second part, which will appear in the next issue of *Scientific American,* tells the story of the AIDS virus.

Early Skepticism

The prologue to the discovery of the first human retrovirus is a history of skepticism: first, doubt that infectious agents would prove to be an important source of cancer and then doubt that retroviruses would ever be found in human beings. It was established quite early that retroviruses do have the capacity to cause tumors, at least in the laboratory. The first retrovirus was isolated in 1910 by Peyton Rous of the Rockefeller Institute for Medical Research. He showed that the virus—now called avian sarcoma virus—can induce tumors of muscle, bone and blood-vessel tissue in chickens.

Rous's results met with such widespread disbelief that he gave up retrovirus research, and the field passed into a dormancy that lasted until the 1950's. In that decade and the next, Ludwik Gross of the Mount Sinai School of Medicine as well as other investigators found retroviruses that cause tumors in mice, chickens and other species. Yet most biologists continued to doubt that infectious agents had a significant role in the transmission of cancer outside the laboratory. After all, the animals studied in the retrovirus work were, by and large, inbred laboratory strains. Moreover, many of the infections were congenital, which seemed unlikely to be an important mode of transmission in human beings.

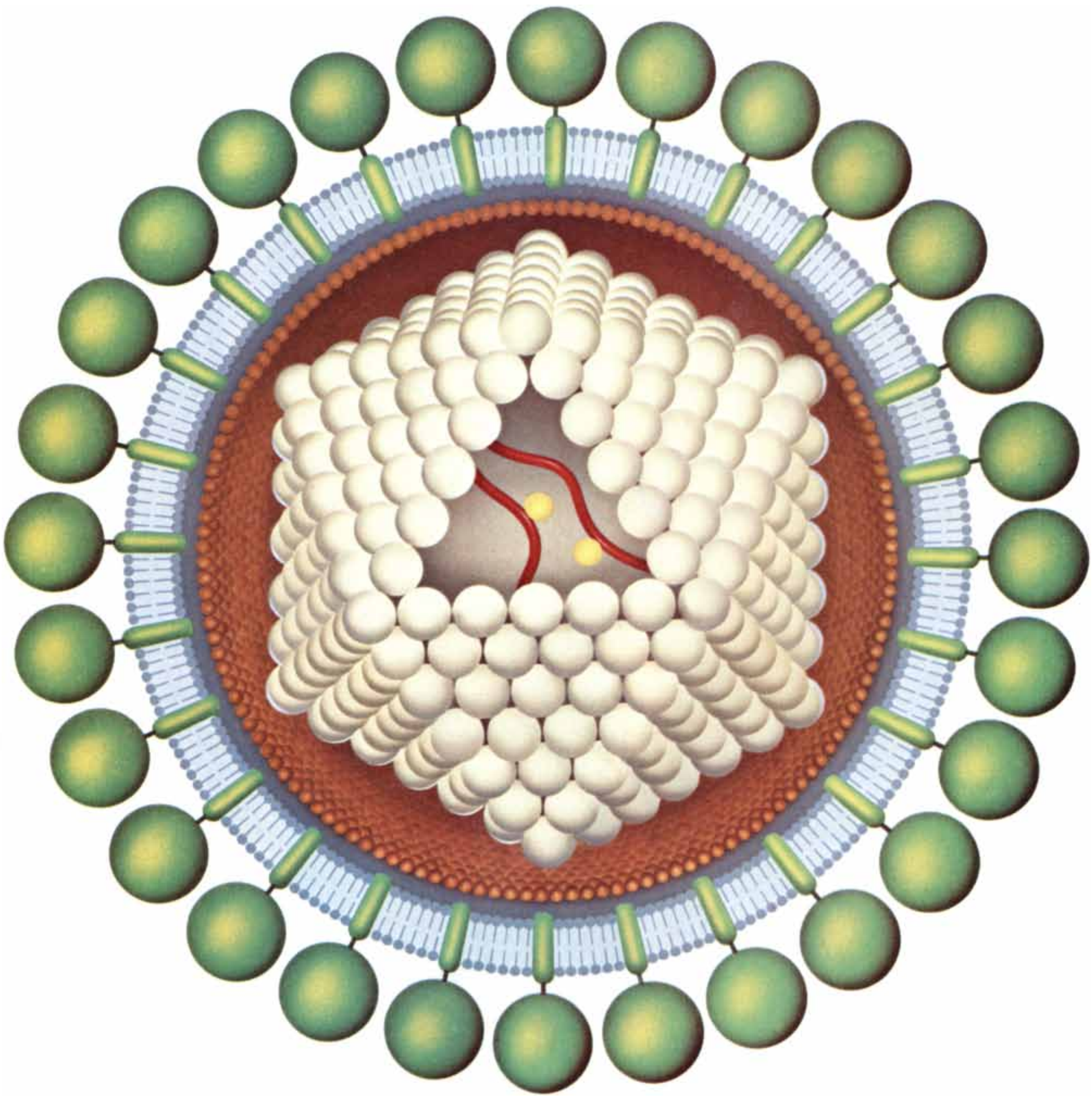
Hence in 1960 it was possible to argue, plausibly, that retroviruses were little more than laboratory curiosities. All that changed in the early 1960's when William Jarrett of the University of Glasgow discovered feline leukemia virus (FeLV). Jarrett, who had been trained as a veterinarian, demonstrated that FeLV was capable of causing not only malignancies of blood cells but also aplasias (insufficient growth of affected cells) and an immune deficiency similar to the one later observed in AIDS patients. Equally important, Jarrett (with his brother and colleague Oswald Jarrett, Myron Essex of Harvard University and William D. Hardy, Jr., of the Memorial Sloan-Kettering Cancer Center) showed that the diseases caused by FeLV were communicated among unrelated cats in the natural setting of the household.

With the discovery of FeLV, retroviruses transcended the status of laboratory curio. Yet there was no rush to find analogous infectious agents in human beings. One reason was the attention then being given to endogenous retroviruses. Endogenous retroviruses are viruses whose genetic material is found in the chromosomes of many animal species, presumably as the remnant of an ancient infection. In some instances such endogenous DNA sequences are capable of giving rise to functional, infectious virus particles. By and large, however, they are transmitted to other members of the species through ordinary Mendelian inheritance: in the sperm or egg. In 1969 Robert J. Huebner and George J. Todaro of the National Cancer Institute proposed that activation of these normally silent endogenous sequences by carcinogens was the mechanism of all malignancy.

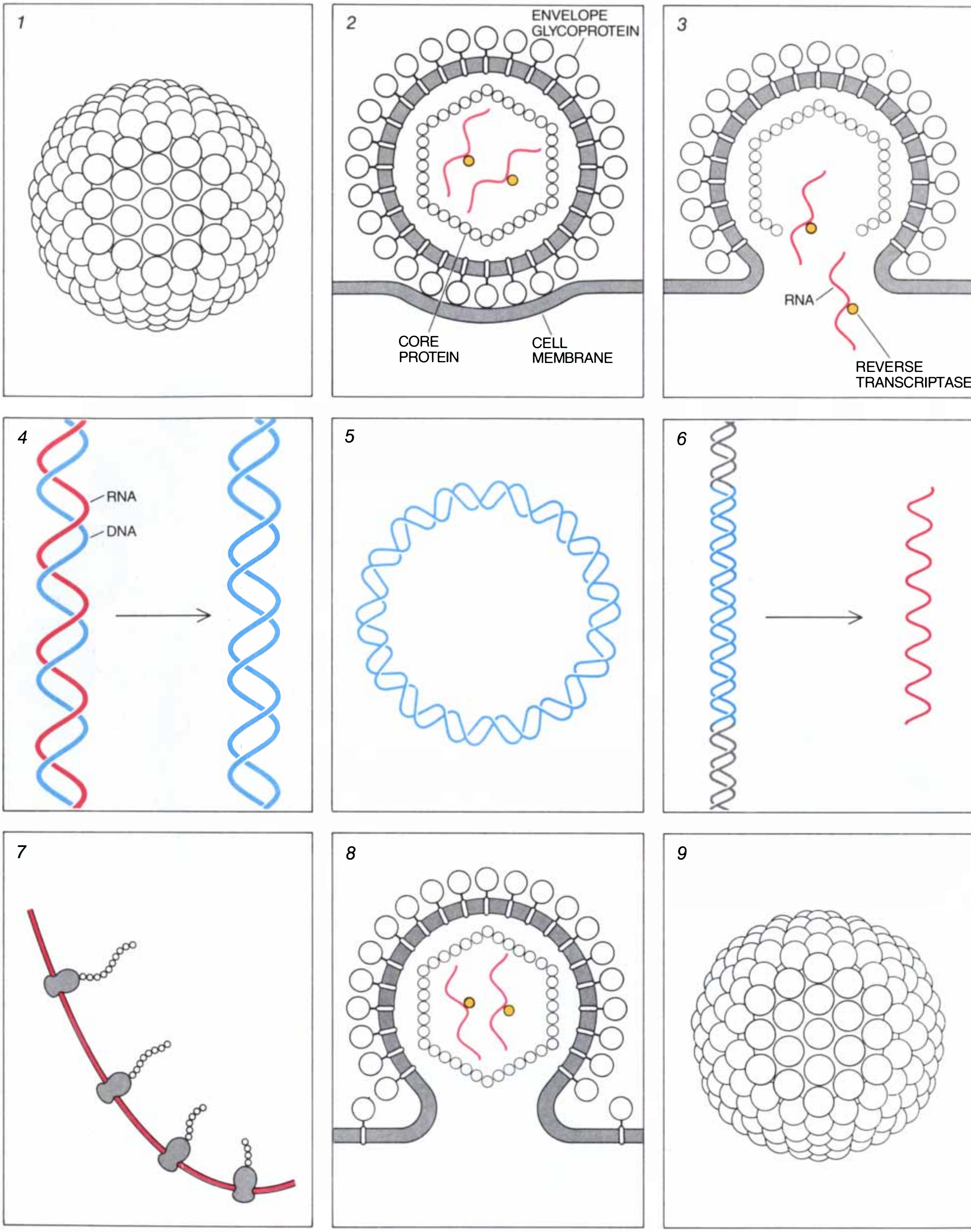
Ironically, it has since turned out that all the diseases known to be caused in nature by retroviruses are due to exogenous viruses. The endogenous viruses are currently regarded as an evolutionary puzzle having little clinical relevance. In the late 1960's, however, exogenous viruses were generally overlooked in the excitement generated by their ubiquitous endogenous relatives.

Reverse Transcription

The seductions of the Huebner-Todaro theory were not the only problem confronting those interested in ex-



PORTRAIT OF A VIRUS reveals much of what is known about the structure of the first human retrovirus: human *T*-cell lymphotropic virus-I (HTLV-I). The illustration shows the virion, or virus particle, in cross section. The particle is about 1,000 angstrom units across (roughly one ten-thousandth of a millimeter). Its outer envelope is a double layer of lipid (fatty) material that is penetrated by proteins (*green*). The envelope covers a core containing several types of proteins (*light brown and ivory*). The core also includes two molecules of RNA (*red*). The RNA comprises the genetic information that is necessary for the virus to synthesize its components and thereby reproduce itself. Bound to the RNA are several copies of an enzyme called reverse transcriptase (*yellow*). Reverse transcriptase exploits the viral RNA as a template for assembling a corresponding double-strand molecule of DNA.



LIFE CYCLE OF A RETROVIRUS includes the integration of the viral genome (the full complement of genetic material) into the DNA of the host cell. The virus particle (*1*) interacts with the membrane of the host cell (*2*), which in the case of HTLV-I is a white blood cell called a *T* lymphocyte. The membrane of the virus fuses with the cell's outer membrane (*3*), releasing the contents of the virion into the cytoplasm. In the cytoplasm reverse transcriptase makes a single strand of DNA corresponding to the viral RNA; then (as the RNA is degraded) a second strand of DNA is made (*4*). The double-strand DNA migrates to the cell nucleus, where it forms a circular structure (*5*) and inserts itself randomly among the host's chromosomes (*6*). Later the viral DNA can be transcribed into RNA, which is translated into protein on cellular ribosomes in the cytoplasm (*7*). Newly made proteins and viral RNA assemble and bud outward (*8*), yielding a new virion that incorporates lipid material from the cell's outer membrane (*9*).

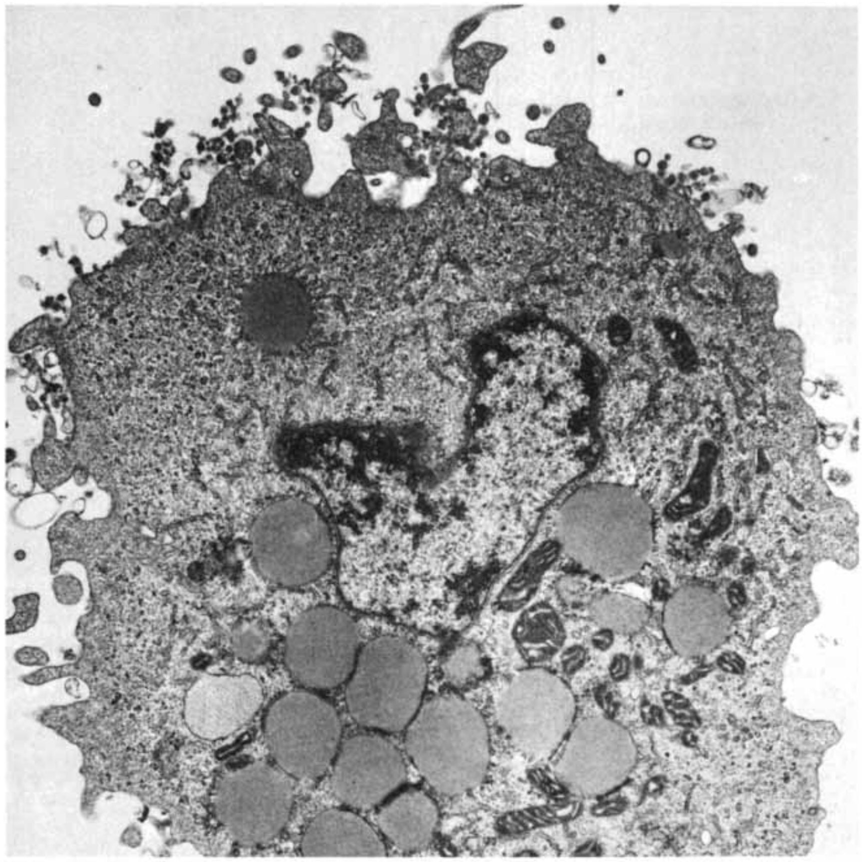
ogenous viruses and cancer in 1969. There was also a conceptual difficulty in understanding how the genome of the virus, which was known to consist of RNA, could interact with the genes of the host cell to induce a tumor. Howard M. Temin of the University of Wisconsin Medical School had proposed that the life cycle of an exogenous retrovirus includes an intermediate DNA phase called a provirus, but the details were hazy.

Temin's discovery of reverse transcriptase (a finding made independently and simultaneously by David Baltimore of the Massachusetts Institute of Technology) provided a concrete basis for the provirus theory, and the details of the life cycle quickly fell into place [see "RNA-directed DNA Synthesis," by Howard M. Temin; SCIENTIFIC AMERICAN, January, 1972]. It became clear that when a retrovirus infects a cell, reverse transcriptase synthesizes a DNA molecule corresponding to the code carried in the viral RNA. The DNA can make its way to the nucleus, where it integrates itself among the genes of the host. Later, through interaction with the host's genes, the provirus may initiate a tumor. Since the viral DNA carries the information needed to synthesize the components of the virus, it can also be activated to form new virus particles.

In 1970, when Temin and Baltimore came on reverse transcriptase, I was studying DNA polymerases in blood cells. DNA polymerases are enzymes that assemble DNA; reverse transcriptase is a member of this group, albeit an unusual one. Under the influence of Temin's ideas I decided to search for reverse transcriptase in human leukemic cells, hoping to find a retrovirus there.

In doing so I was gainsaying accepted wisdom. The animal leukemia viruses (by then well known in chickens and mice as well as cats) undergo extensive viral replication before a tumor is initiated; the new virus particles are readily visualized in the electron microscope. Careful electron microscopy of human leukemic cells had yielded no such images, leading most investigators to conclude no human retrovirus exists.

Yet it seemed to me—and, independently, to Sol Spiegelman of Columbia University—that the search was not in fact pointless. Perhaps human retroviruses exploit a different cancer-causing mechanism, which does not entail extensive viral replication. If so, electron microscopy (a cumbersome tool for the purpose) would never detect the pathogen. A more sensitive assay, though, might, and fortunately one was at hand in reverse transcriptase.



NEW VIRUS PARTICLES bud from a *T* lymphocyte (enlarged 13,000 diameters) in laboratory culture. The particles appear as small, dark circles at the periphery of the cell. The irregular shape at the center is the nucleus; the rounded areas are lipid droplets. The image is unusual, since HTLV-I replicates only during a brief part of the *T*-cell life cycle. The electron micrograph was made by Bernhard Kramarsky of Electro-Nucleonics, Inc.

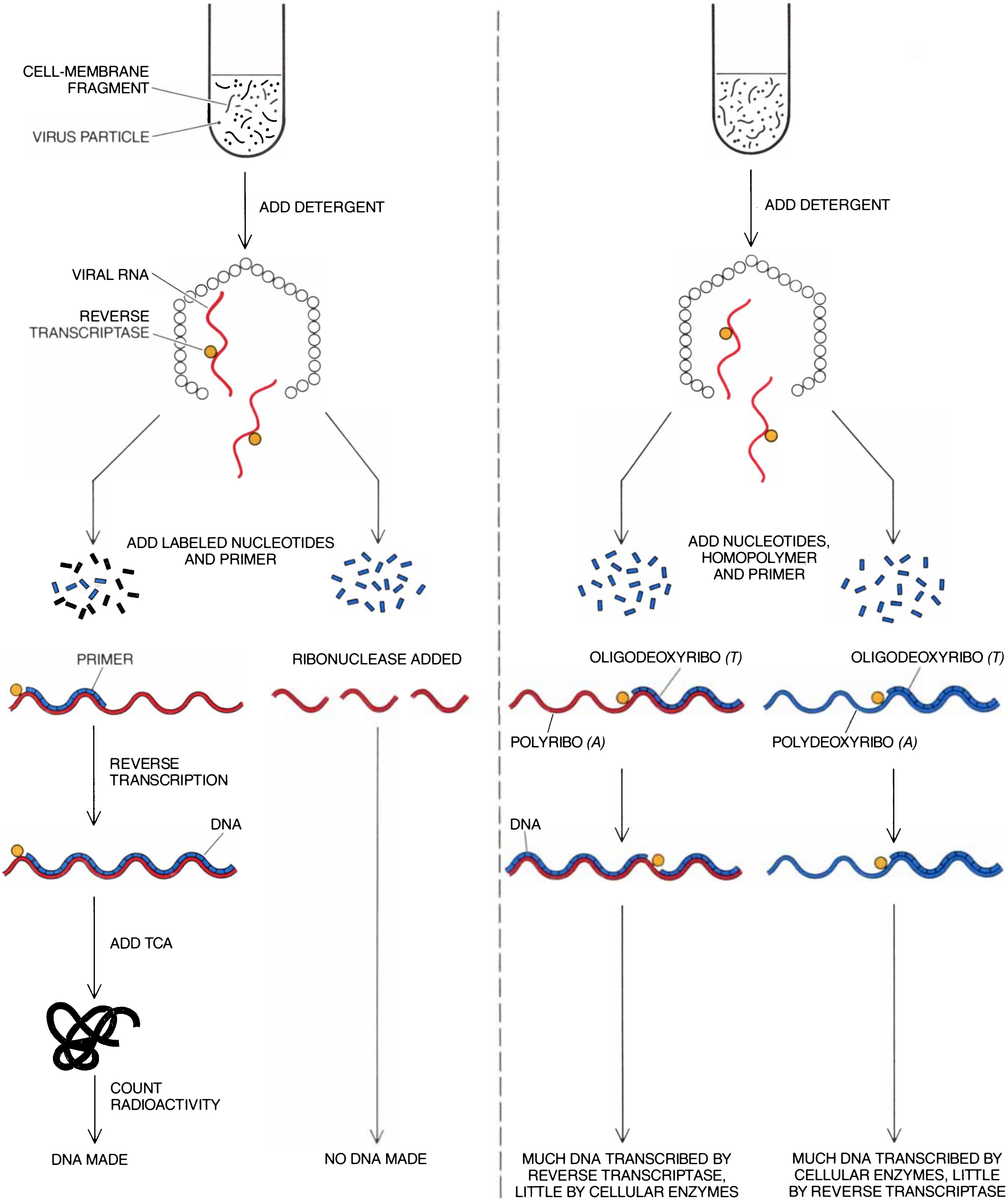
Since reverse transcriptase is unique to retroviruses, finding it in tumor cells would show that such a virus was there. Furthermore, a biochemical assay for the activity of the enzyme in assembling DNA could be made far more specific and sensitive than microscopy.

The Virtues of Culture

Between 1970 and 1975 Spiegelman's group and my own labored to refine the reverse-transcriptase assay until it was several orders of magnitude more sensitive than electron microscopy [*see illustration on next page*]. Using such assays, my colleagues and I obtained some tantalizing but still inconclusive results. From the leukemic cells of a few patients we purified DNA polymerases that seemed to have all the properties of reverse transcriptase. There was, however, an ambiguity: the enzymes might have been unusual cellular polymerases detectable only because their numbers are increased in diseased cells. To demonstrate their origin unequivocally we needed a much larger supply of the putative virus, which in turn required us to grow the infected cells.

Now, human blood cells—normal or cancerous—are difficult to grow in the laboratory. Indeed, in the early 1970's no laboratory cell line existed that was suitable for our purposes. By that time, however, proteins called growth factors had been discovered that were useful for making other recalcitrant cells grow in the laboratory. It seemed reasonable to think growth factors could be the key to growing white blood cells. Accordingly, a search was begun for growth factors suitable for our purposes.

A step in that direction had been taken during the 1960's by Peter C. Nowell of the University of Pennsylvania School of Medicine. Nowell found that a protein called phytohemagglutinin (PHA), derived from plants, could induce certain white blood cells to grow larger, become active and divide once or twice. Nowell's cells were later shown to be *T* cells, and his experiment amounted to a short-term culture. Picking up where Nowell had left off, in 1976 my colleagues Doris Morgan and Francis



REVERSE-TRANSCRIPTASE ASSAY is based on identifying the enzyme's capacity to assemble DNA from an RNA template. In its original form (*left*) the assay relied on endogenous viral RNA as the template. Infected cells were centrifuged to separate the virus particles (and pieces of cell membrane) from other material. A mild detergent broke up the particle, releasing the reverse transcriptase. Radioactively labeled nucleotides (the subunits of DNA) as well as a specific piece of DNA that serves as a "primer" for DNA assembly were added to the solution containing the enzyme. The reverse transcriptase synthesized DNA, which was precipitated by trichloroacetic acid (TCA). Measuring the radioactivity of the solid showed that the labeled nucleotides had indeed been incorporated into DNA. If ribonuclease (an enzyme that breaks up RNA) is added to the solution, no DNA is made, showing that the process depends on an RNA template. Later the assay was modified (*right*) by the use of synthetic templates, making it possible to distinguish conclusively between the activity of reverse transcriptase and the activity of related cellular enzymes that exploit DNA as a template. If the template called polyribo(*A*)—a form of RNA—is used, the viral enzyme makes much DNA but cellular enzymes make little. If the template called polydeoxyribo(*A*)—a form of DNA—is used, the situation is reversed.

Ruscetti and I found that after PHA stimulation some *T* cells did indeed release a growth factor. We named the new substance *T*-cell growth factor, or TCGF; it is now generally called interleukin-2, or IL-2.

Not only did the activated *T* cells secrete IL-2 but also they developed receptor molecules on their surface for the same protein. When the growth factor bound to its receptor, the cells began to divide. We had found a way to culture *T* cells: after accumulating enough IL-2 we could add it to the prestimulated cells and thereby maintain their growth for long periods. As it happens, this method parallels events in the normal immune system. *T* cells, which stem from a precursor in bone marrow, migrate to the thymus gland to mature. There they become differentiated into two classes—*T*4 and *T*8—with different marker molecules on their surface and different immune functions. In response to infection some of these mature *T* cells are first activated by the protein called IL-1 and then stimulated to divide by IL-2.

In the late 1970's these pathways were but dimly perceived. Not until later did work by immunologists—notably Kendall A. Smith and his colleagues at the Dartmouth Medical School and Hans Wigzell of the Karolinska Institute in Stockholm—demonstrate the full significance of IL-2 for the immune system. In any event, our primary interest was not in the immune reponse as such but in growing *T* cells, and for that purpose IL-2 was very effective.

Indeed, Bernard Poiesz, a postdoctoral fellow in my group, found that some leukemic human *T* cells could be grown with IL-2 without prior activation by PHA. The function of PHA stimulation was to generate the IL-2 receptors, which normal *T* cells lack. Apparently the leukemic *T* cells had the IL-2 receptor already and so did not require activation. As I shall describe, this surprising finding has turned out to be of significance for how HTLV-I causes cancer. Further work on the leukemic *T* cells revealed that they were relatively mature and had the *T*4 marker molecule.

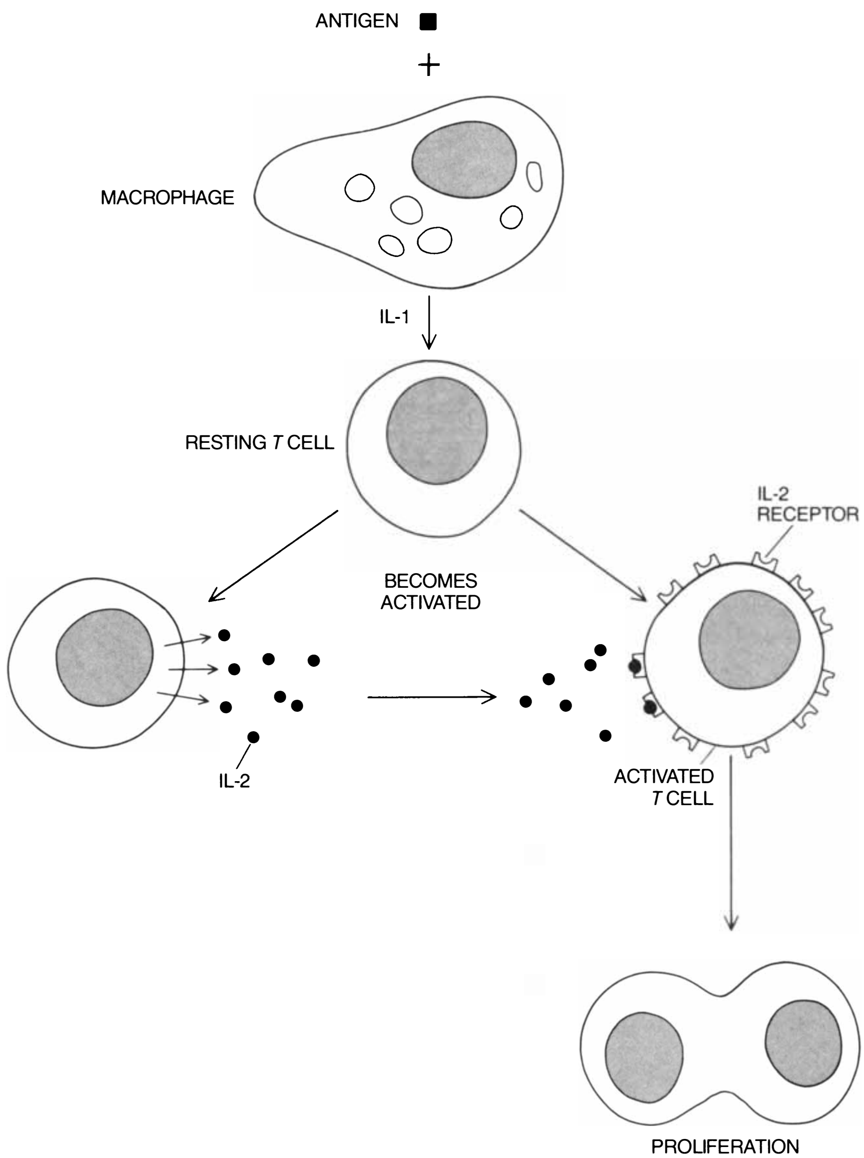
The First Isolates

It was from such malignant cells, grown with IL-2, that we isolated the first examples of HTLV-I in 1978–79 from the cells of two leukemia patients. My colleagues and I isolated the virus, characterized it and showed it was specifically a human virus; our results were published in 1980 and early 1981. Later we isolated many other examples. My colleague Marvin Reitz then showed that these viruses were not closely related to previously described animal viruses. Equally important, he showed that our isolates were not endogenous but exogenous and therefore the same type of virus that causes disease in animals.

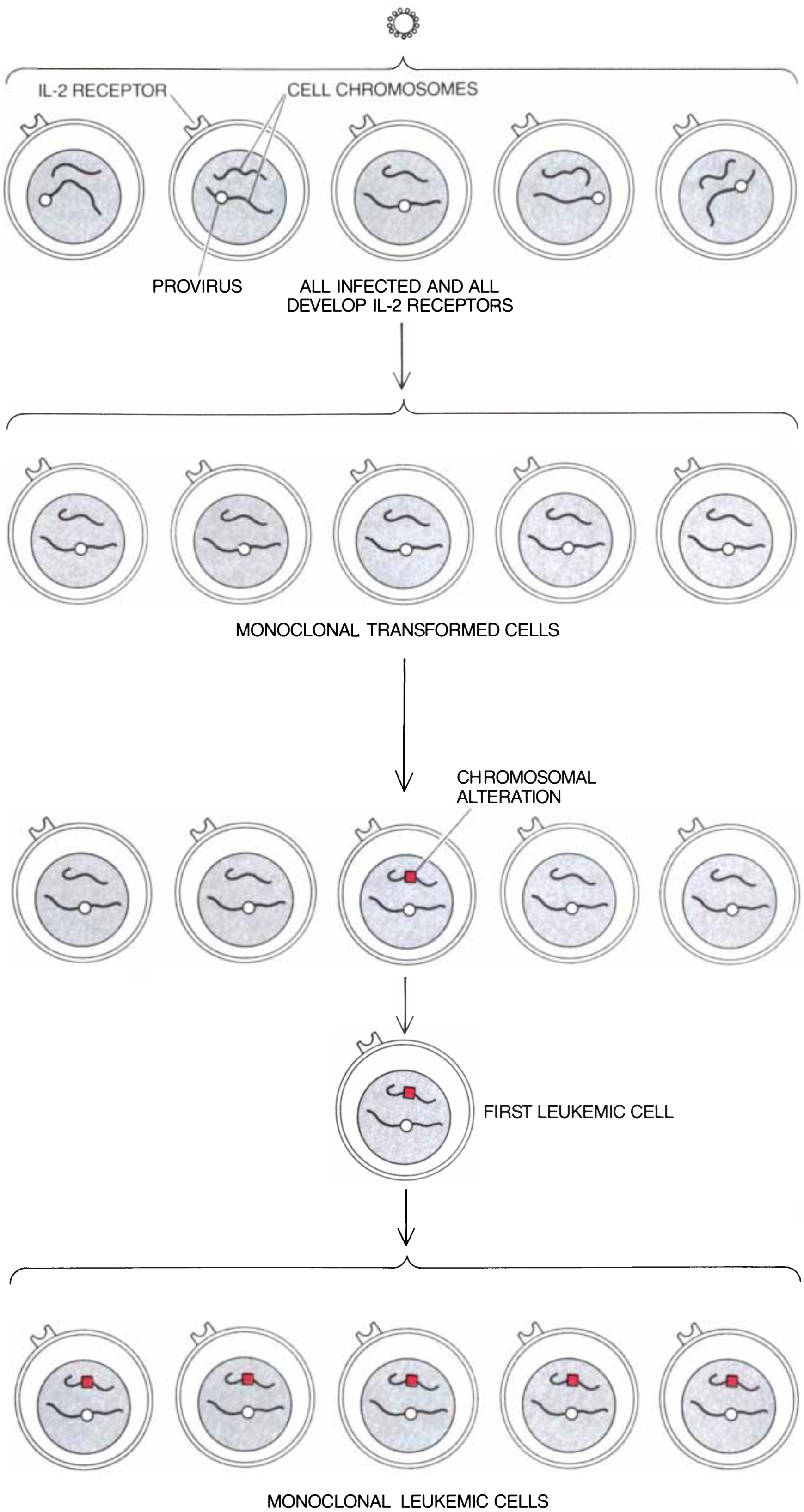
There was, however, some confusion about the disease HTLV-I was associated with. This was bound to be the case, since the origin of *T*-cell leukemias was not well understood, and so clinicians had been forced to rely on symptoms to categorize the forms of the disease. The patients from whom we first isolated HTLV-I had malignancies of mature *T*4 cells accompanied by skin abnormalities, which result from infiltration of the skin by malignant blood cells.

Such a clinical picture has been called mycosis fungoides or Sézary *T*-cell leukemia. It was evident from our early studies, however, that HTLV-I could be detected in only a small fraction of patients with mycosis fungoides or Sézary syndrome. Conversely, many leukemia patients who were positive for HTLV-I lacked skin abnormalities and so did not fit the clinical definition of these conditions.

Some light was cast into this ob-



***T* CELLS ARE ACTIVATED in response to infection. The process begins when a white blood cell called a macrophage encounters an antigen—a surface protein from an invading organism. The macrophage secretes a protein called IL-1. When IL-1 reaches resting *T* cells, they secrete a second protein, called IL-2, and develop IL-2 receptors on their surface. The binding of IL-2 to its receptors induces the *T* cell to divide and mature (assume the functional characteristics needed for its complex role in the immune response). The discovery that IL-2 (originally called *T*-cell growth factor, or TCGF) could be used to grow *T* cells in the laboratory was a crucial step in the author's isolation of HTLV-I.**



HTLV-I CAUSES LEUKEMIA by a multistep process. When the virus infects a group of *T* cells, all the cells develop IL-2 receptors, and the provirus (the viral DNA integrated into the cellular genome) is found in each cell. The provirus, however, is found in a different position in each cell. Some of the infected cells are transformed into a precancerous state; some of the transformed cells then give rise to clones of descendants, each of which contains an identical copy of the progenitor's DNA, including the provirus. In one cell of such a clone a further chromosomal alteration may lead to the development of the final cancerous state. All the leukemic cells are descended from this single cancerous ancestor.

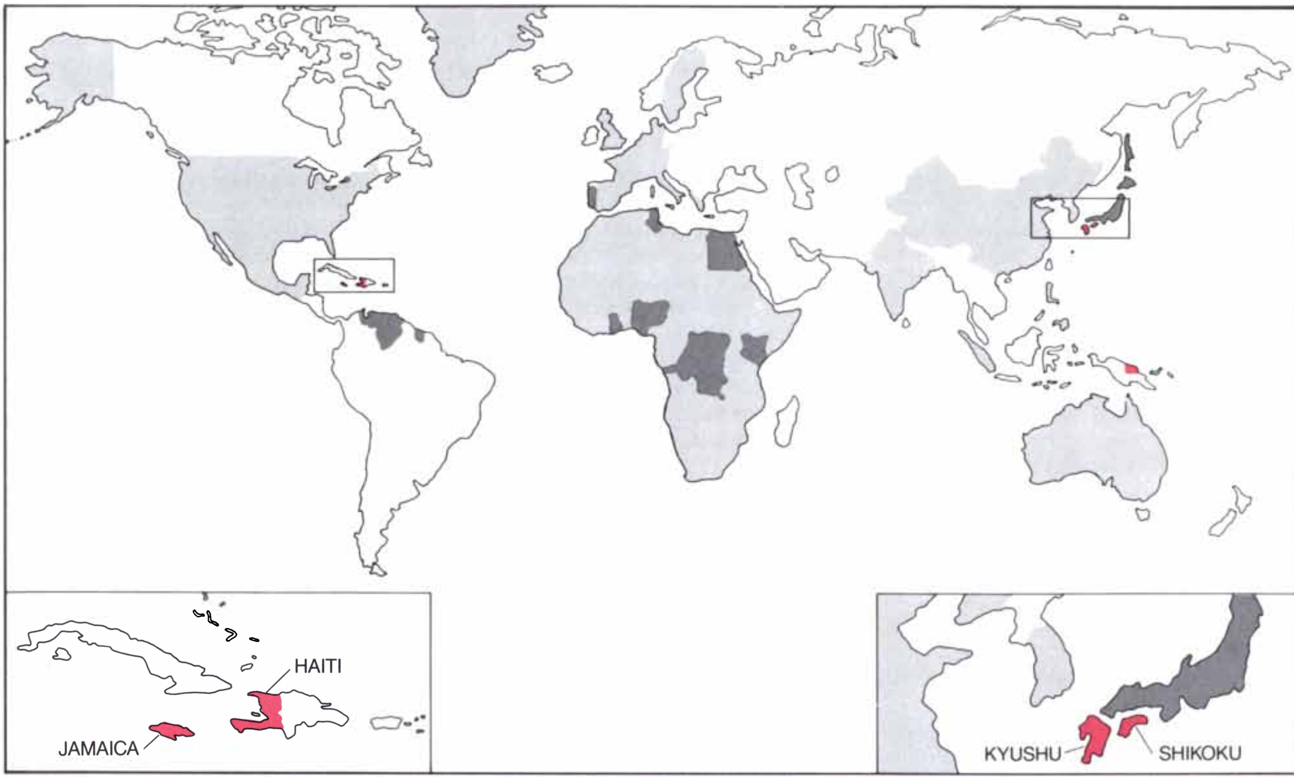
scurity when we learned of a disease called adult *T*-cell leukemia (ATL), first described in 1977–78 by Kiyoshi Takatsuki of Kyoto University. ATL resembles mycosis fungoides and Sézary syndrome, but there are some significant differences. For one thing, the skin is not always involved. In addition, ATL is more aggressive than the other two syndromes (median survival from the time of diagnosis is only three or four months). Its victims—although still adults—are generally younger, and the malignant *T* cells in the blood are often accompanied by hypercalcemia (an increase in the blood's calcium content). The hypercalcemia alone can be fatal, but more frequently death results from explosive proliferation of the leukemic cells or from opportunistic infections.

Emergent Geography

Takatsuki and his co-workers noted that ATL is heavily concentrated in Kyushu and Shikoku, the southernmost major islands of Japan. Such clustering suggested the disease might be caused by an infectious agent. Since the symptoms resembled those of the cases that had yielded HTLV-I, we began collaborating with Yohei Ito of Kyoto University to find out whether ATL patients were also harboring the virus. They were. Blood serum from all ATL patients tested contained antibodies that reacted with HTLV-I. What is more, monoclonal antibodies (preparations of specific antibody that react with only one protein) made to HTLV-I proteins reacted with ATL cells but not with normal cells. It looked as though HTLV-I was the cause of ATL.

Supporting evidence for that conclusion was offered at the virus workshop in Kyoto where, in March of 1981, I presented our early results. Yorio Hinuma of Kyoto University described a line of ATL cells developed by Isao Miyoshi of Kochi University that, when grown in the laboratory, released retrovirus particles. All available data indicated that the virus coming from Miyoshi's cells was identical with HTLV-I.

Not long afterward the point was clinched. Mitsuaki Yoshida of the Tokyo National Cancer Institute worked out the sequence of nucleotide bases in the RNA genome of the Japanese isolate and compared it with the sequence of HTLV-I. The overlap was sufficient to make clear that the Japanese and American isolates were very closely related strains of a single virus. The first independent confirmation of the presence of HTLV-I in the U.S. was later provided by Barton Haynes and



GLOBAL PATTERN OF HTLV-I INFECTION yields clues to the origin of the virus. The main endemic areas of the virus (*color*) are the Caribbean basin and southwestern Japan. Another possible endemic area, as yet unconfirmed, is in New Guinea. Areas with intermediate levels of infection (*dark gray*) include parts of equatorial Africa and Latin America. Low levels of infection (*light gray*) are present in other areas. Epidemiological data are lacking, however, for much of the world (*white*). The finding of viruses closely related to HTLV-I in certain species of African monkeys suggests the virus may have originated in Africa and been transported to the rest of the world through commerce, in particular the slave trade. The data on the map were collected by William A. Blattner in collaboration with Carl Saxinger, Marjorie Robert-Guroff and the author (all of the National Cancer Institute).

Dani Bolognesi of the Duke University Medical Center, who isolated the virus from a Japanese-American suffering from a *T*-cell leukemia.

The clustering of ATL in southern Japan had been significant in helping to link HTLV-I to a specific disease, and my co-workers and I wanted very much to be able to identify similar patterns elsewhere. Until 1980, however, our epidemiological work was severely hampered by the absence of apparent disease clusters in the U.S. The only salient epidemiological fact was that most of our patients were blacks born in the U.S., the Caribbean countries or South America. Then in 1981 Daniel Catovsky of Hammersmith Hospital in London pointed out a concentration there of *T*-cell leukemias among Caribbean-born blacks. Their clinical features were remarkably similar to those of the U.S. and Japanese cases that had proved positive for HTLV-I, and as we by now expected HTLV-I was found in all of them.

Catovsky's insight provided one of the keys that unlocked the international epidemiology of HTLV-I. With the formidable cooperation of William A. Blattner of the National Cancer Institute and his associates, my colleagues Carl Saxinger and Marjorie Robert-Guroff and I began surveying black populations in the U.S., the Caribbean, South America and Africa for HTLV-I infection. It was found that HTLV-I is endemic not only to the southern islands of Japan but also to parts of the U.S., most of the Caribbean, northern South America and, in particular, to Africa.

Commerce and Cancer

What could tie these disparate regions together? The answer, remarkably enough, appears to be the slave trade, but that hypothesis did not emerge directly. Miyoshi discovered that some Japanese macaques had antibodies to HTLV-I, and he speculated that the virus had infected people in Japan from these monkeys. That early proposal was shown to be false: the Japanese macaque virus differs enough from HTLV-I to rule out direct transmission. In confirming the results obtained by Miyoshi, however, Gebhard Hunsman of the University of Göttingen and my group found that many species of African monkeys also have antibodies that react with HTLV-I. The viruses subsequently isolated from those monkeys were also related to but distinct from HTLV-I.

Some of the African viruses—notably those from African green monkeys and from chimpanzees—proved to be much more closely related to HTLV-I than the Japanese macaque virus is. The resemblance was sufficient to suggest there is a close connection between the infection of those species and that of human beings. On the basis of such results (along with the epidemiology and some historical information) I proposed the following hypothesis. HTLV-I originated in Africa, where it infected many species of Old World primates, including human beings. It reached the Americas along with the slave trade.

Curiously, it may well have arrived in Japan the same way. In the 16th century Portuguese traders traveled to Japan and stayed specifically in the islands where HTLV-I is now endemic. Along with them they brought both African slaves and monkeys, as contemporary Japanese works of art show, and either one or the other may have carried the virus. This hypothesis has recently been challenged by the finding that HTLV-I infection is common among the Ainu people living on Hokkaido, Japan's northernmost major island—an area where the Portuguese did not go. For the time being, however, it remains a plausible explanation of the global pattern of spread.

The epidemiological findings about HTLV-I clarified not only global patterns but also patterns in smaller areas, and even the methods by which the virus is transmitted from one person to another. The basic linkage of HTLV-I to ATL was relatively uncomplicated because the virus is more localized than some others and because the correspondence between virus and disease is marked: where clusters of ATL are found, the virus is prevalent; the converse is also true. Within small geographic areas such as townships, however, the prevalence of infection varies greatly.

The great variation in prevalence within small areas (along with other data) shows that HTLV-I is not casually transmitted. Several routes of infection have been established, all requiring close exchanges. The virus can be transmitted in contaminated blood transfusions or among drug addicts sharing a needle. It can be transmitted by sexual contact, either homosexual or heterosexual. A fetus can be infected in the womb if its mother harbors the virus. Recent observations from Japan suggest that infants can ingest HTLV-I in their mother's milk. Even more recent findings indicate that the virus can also be carried by mosquitoes. Once the virus has been passed, there follows a latency that can be as long as 40 years (if infection occurs in infancy) or as short as a few years (if it occurs in adulthood).

The Direct Cause

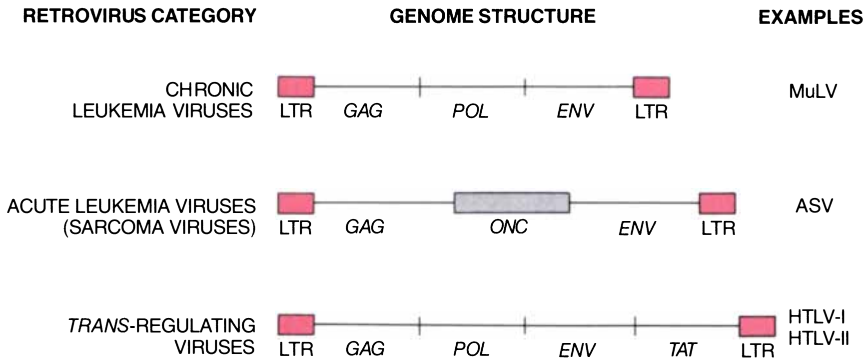
What is the relation between the original infection and the tumor that may ensue as much as 40 years later? Several lines of evidence suggest that the virus's role is quite direct. One type of evidence is epidemiologic. Infected infants born in the endemic area of southern Japan have the same chance of developing ATL whether they spend the rest of their lives there or move to another part of the world at an early age; the same is true for other endemic areas. It appears that the virus can initiate the chain of events leading to a tumor on its own, and that no environmental factors present in the endemic area are needed.

The second line of evidence comes from molecular virology. When HTLV-I infects *T* cells in the laboratory, the provirus is found integrated randomly in the cellular genome: in one cell it may be on chromosome 5, in another on chromosome 8, in a third on chromosome 15. Yet when the cells of an ATL patient are examined, the viral sequences are found in the same place in every cell of the tumor. Such identity implies that the tumor is a clone: all its cells are the progeny of a single ancestor and as such contain copies of the progenitor's genome. It also implies that the infection preceded the origin of the tumor, because if the virus had entered the cells of an existing multicellular tumor, the viral sequences would be found in a different place in each cell of the tumor.

PORTUGUESE TRADERS IN JAPAN may have brought HTLV-I with them in their African slaves or in monkeys. During the 16th century Portuguese seafarers made many visits to Japan. The illustration shows a contemporary Japanese depiction of one such visit. The figure at the left is an African holding a canopy over two Portuguese merchants engaged in animated conversation. Contact with the Portuguese was concentrated in the southernmost islands of Japan. In 1978 a new type of *T*-cell malignancy was identified in the same region of Japan. The disease—adult *T*-cell leukemia, or ATL—has been strongly linked to HTLV-I.

That is not all. When Miyoshi cultured ATL cells with human umbilical-cord blood, *T* cells from the (normal) baby's blood were transformed into an immortal, precancerous condition. It seemed likely that virus particles escaping from the leukemic cells had transformed the normal ones. That explanation was later unequivocally confirmed in my laboratory in experiments begun by Mika Popovic and continued by Zaki Salahuddin and Philip Markham. They showed that many isolates of HTLV-I can transform *T*4 cells in from five to seven weeks; the provirus of the transformed cells is integrated in a pattern like the one seen in ATL cells. The circle was closed by Miyoshi, who took laboratory-transformed cells, put them in hamsters and showed they can cause malignancies.

Linking HTLV-I directly to a cancer was exciting, but it was hardly an end point. On the contrary, it provided the starting point for the next goal, which was to identify the molecular mechanism by which the virus causes leukemia. Although that project is not yet complete, some of the outlines of the disease process have emerged from the work done so far. One of the central insights has come from examining the pattern of integration of the provirus in the cellular genome (a pattern worked out by my colleagues Flossie Wong-Staal, Beatrice Hahn and Vittorio Manzari and, independently, by Motoharu Seiki of the Tokyo Cancer Institute working with Yoshida). As I mentioned above, in each cell of a particular tumor the provirus is integrated in a clonal fashion. From one ATL patient to another, however, there is no uniformity: each has the viral sequences in a different location.



LEUKEMIA-CAUSING RETROVIRUSES can be divided into three groups according to their genetic structure and pathogenic mechanism. The basic retroviral genome includes DNA sequences that code for the core proteins, reverse transcriptase and the viral envelope proteins; they are known as *gag*, *pol* and *env* respectively. The chronic leukemia viruses, such as mouse leukemia virus (MuLV), have only those three types. The acute leukemia viruses, or sarcoma viruses, have another DNA sequence: a cellular gene known as an *onc* gene, which is involved in regulating cell growth. Among the sarcoma viruses is the avian sarcoma virus (ASV). The third group, the *trans* regulating viruses, includes HTLV-I and its relative HTLV-II; it is defined by the presence of a gene called *tat*.

A Novel Mechanism

The combination of clonality with randomness of integration from tumor to tumor immediately indicated that HTLV-I must cause leukemia by a new mechanism. Before HTLV-I only two mechanisms were known whereby a retrovirus can cause leukemia, each defining a group of viruses. One mechanism entails the presence in the viral genome of an *onc* gene, a cellular gene involved in the regulation of growth. Expression of the *onc* gene transforms every infected cell, giving rise to a rapidly developing cancer, which (since it develops from many different progenitors) is not monoclonal. That mechanism is shared by a variety of acute leukemia viruses and sarcoma viruses (including the avian sarcoma virus isolated by Rous). The genetic event that leads to the requisite combination of cellular and viral DNA is rare, however, and these viruses are of little general consequence.

Much commoner are the chronic leukemia viruses, including avian leukemia virus (ALV) and mouse leukemia virus (MuLV). They lack an *onc* gene and the disease they cause is much more like the natural human malignancies: only a minority of infected animals develop leukemia; the leukemias appear only after a long latency; the tumors are clonal. But there is a critical difference between their mechanism and that of HTLV-I. For these viruses to cause disease, the provirus must integrate in a specific spot, enabling regulatory sequences in the viral DNA to interact with nearby cellular genes that promote cellular proliferation. HTLV-I, on the other hand, is apparently pathogenic no matter where in the genome its provirus integrates.

In addition to guiding ongoing research, the contrast between the chronic leukemia viruses and HTLV-I helped to clear up an old puzzle. As I noted above, much of the doubt that human retroviruses would ever be found stemmed from the fact that electron microscopy failed to detect virus particles in human tumor cells. As the mechanism of HTLV-I unfolded, a possible reason for the low level of virus in the cancer cells became clear. Since HTLV-I can trigger the chain of events leading to a tumor by integrating anywhere in the cellular genome, its mechanism requires few virus particles. In contrast, the mechanism of the chronic leukemia viruses depends on integration in a specific location, and extensive replication may be a way of helping to ensure that at least one copy of the provirus reaches the right spot.

Action at a Distance

Clearly there are significant differences between the modes of action of the two types of virus. Yet it is generally assumed that both mechanisms ultimately affect the activity of cellular genes controlling proliferation. The provirus of the chronic leukemia viruses does so by virtue of its proximity to those cellular sequences. The HTLV-I provirus, on the other hand, may be integrated anywhere in the cell. If HTLV-I is to turn on the growth-promoting genes, it must do so by acting at a distance. Molecular biologists call such mechanisms *trans*-acting. What is the *trans*acting mechanism of HTLV-I?

The beginning of an answer came when Seiki and Yoshida worked out the nucleotide sequence of the provirus and found a region, novel among known retroviruses, that they called *X*. They noted that *X* could potentially encode four proteins. Wong-Staal and my colleague George Shaw, in collaboration with William Haseltine of the Dana-Farber Cancer Institute, soon identified a segment of *X* that turned out to be the gene for one of the four. The gene, common to all isolates of HTLV-I, was called *tat* for *trans*-acting activation. It has since been shown that the protein encoded by the *tat* gene is essential to the transforming effects of the virus in culture.

A clue to what *tat* might be doing came from some early attempts to grow *T* cells. As I have described, we had found that some malignant *T* cells could be grown with IL-2 without any prior activation. Ordinarily IL-2 is in short supply in the body and the genes for the IL-2 receptors are turned off. Our findings, however, suggested that the genes for IL-2 might be turned on and kept on in the malignant cells. In an exciting recent experimental result my colleague Warner Greene found that when the *tat* gene was inserted into human *T* cells, the cells not only made IL-2 but also began to produce IL-2 receptors. It seems probable that early in transformation the *tat* gene induces the production both of IL-2 and of its receptors; the doubly induced cells then undergo abnormal growth.

Yet there must be other events in the pathway leading to malignancy, because by the time a patient has full-blown ATL the leukemic cells neither make nor need IL-2. In addition, at the stage of frank leukemia none of the viral genes, including *tat,* are expressed at detectable levels in the tumor cell. It is likely that after transformation is initiated by *tat* the infected cell proliferates extensively. But a second (and perhaps a third) genetic event must be needed to make the cancerous state final in some of the descendant cells.

Intriguingly, the product of the *tat* gene has a role in viral replication as well as in initiating cancer. At each end of the provirus is a stretch of DNA called the long terminal redundancy, or LTR. The LTR's include sequences that regulate the expression of the viral genes and thereby determine the rate at which the components of the virus (RNA and proteins) are made. By interacting with some of the regulatory sequences, the *tat* protein can quickly increase the level of production of new virus particles. This may be of great benefit to HTLV-I, which replicates only during a restricted portion of the cellular life cycle (when the *T* cells are activated) and probably needs to exploit this brief "window" to the full.

Other Avenues

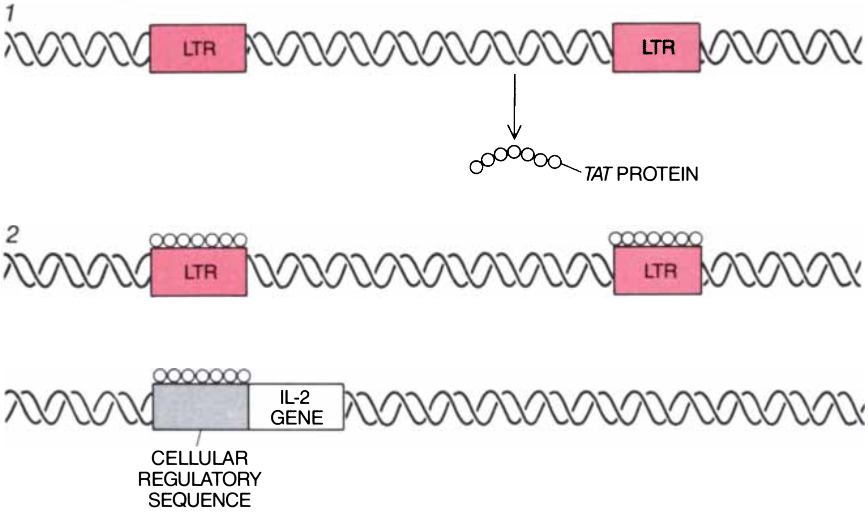
That, for the moment, is a capsule summary of what is known about the mechanisms of HTLV-I on the level of molecules. While work proceeds on the molecular mechanisms, several other avenues of investigation are also being pursued. One concerns the range of diseases with which the virus is associated. In addition to being the direct cause of *T*-cell malignancies in adults, HTLV-I may be an indirect contributing factor in several other pathological conditions.

For example, Essex has suggested that people infected with the virus are prone to other infections, perhaps because some infected *T* cells, although not transformed, are functionally impaired. Such infected "normal" *T* cells may also contribute to leukemias of another type of white blood cell—the *B* lymphocyte—that are observed in some patients from endemic areas of the Caribbean. (The connection between the infected *T* cells and the leukemias of *B* cells may point the way toward a deeper understanding of how leukemia develops.) Finally, recent work has shown that HTLV-I is associated with a neurological disease resembling chronic multiple sclerosis. It seems clear that the overall impact of HTLV-I on public health is just beginning to be recognized.

Work is also proceeding on the second human retrovirus, HTLV-II, which was discovered in 1982 by my group in a collaborative study with David W. Golde of the University of California School of Medicine at Los Angeles. The first isolate came from a line of *T* cells taken from a young white man who had a *T*-cell variant of a disease known as hairy-cell leukemia. HTLV-II differs subtly in form and function from its viral cousin and is associated with less aggressive *T*-cell leukemias. In the most significant ways, however, the two viruses are quite similar (as was demonstrated by Golde and his co-worker Irvin Chen as well as by my colleagues Wong-Staal and Edward Gelman). HTLV-I and HTLV-II have the same overall genomic structure, and there is considerable overlap in nucleotide sequence. They show the same capacity for transforming cells in culture. Moreover, they share the same *trans*acting mechanism, and a *tat* gene product—designated *tat* II—has been found in the new virus.

The diseases caused by HTLV-I and HTLV-II are (except in the endemic areas) relatively rare. This may not always be the case. There is evidence that both viruses are spreading among some populations that are at increased risk of infection, particularly intravenous drug users. Disturbed by such findings, the Red Cross is currently carrying out a survey to determine whether it is necessary to screen donated blood for these viruses, as is already being done for HTLV-III. By the time this article appears a decision will probably have been made; early indications are that it will be positive.

Our growing knowledge of the first human retroviruses is the result of a search that began as early as 1910. In that long history one episode—the isolation of HTLV-I and its linkage to a human cancer—was critical. That accomplishment rested on two preconditions. The first, a sensitive assay for the virus, was provided by the discovery of reverse transcriptase. The second was the establishment of a method for growing *T* cells in the laboratory. Without a method for culturing *T* cells there could have been no test for screening blood, no monoclonal antibodies for epidemiological surveys and no DNA probes for understanding the molecular mechanisms of the virus. These techniques form the basis of a method that took on the utmost significance when the world was struck by the first great pandemic of the second half of the 20th century: AIDS.



***TAT* PROTEIN has crucial roles in the replication of HTLV-I and in the transformation of the host cell into the cancerous state. The *tat* gene lies near one end of the provirus, next to a stretch of DNA called the long terminal redundancy, or LTR (*1*). The paired LTR's regulate the activity of other viral genes. By binding to the LTR's the *tat* protein can help to control the rate of viral replication (*2*). By binding to cellular regulatory sequences (which may control the genes for IL-2 and its receptor) the *tat* protein may induce the host cell to proliferate abnormally, perhaps the first step toward malignancy.**

The next step:

Summary:

Remarkably strong ceramics have been developed over the past few years. They have not been widely used in high-stress environments, however, because of their brittle nature and relatively low toughness. But that is changing as research by GTE reveals new ways to achieve high levels of toughness in a growing range of ceramics.

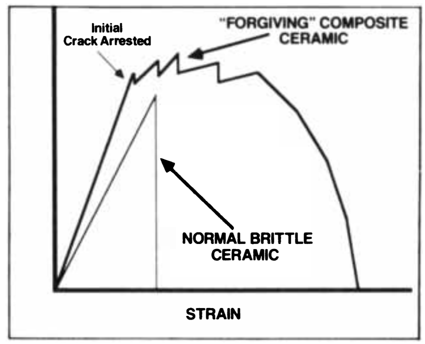
Ceramic science has made enormous strides in recent years. But the same old problem has continued to plague researchers:

Ceramics fail catastrophically, in a brittle manner.

At GTE, we are learning how to produce ceramics approaching metal-like toughness, and with strength levels even higher than presently available.

The science of graceful failure.

In traditional brittle ceramics, increasing stress causes increased strain until a critical value is reached, at which point, a crack begins and propagates catastrophically, and the component and/or part fails.



Traditional ceramics fail catastrophically at critical stress levels. Toughened ceramics (thicker line) retain significant load-bearing capability after initial cracking.

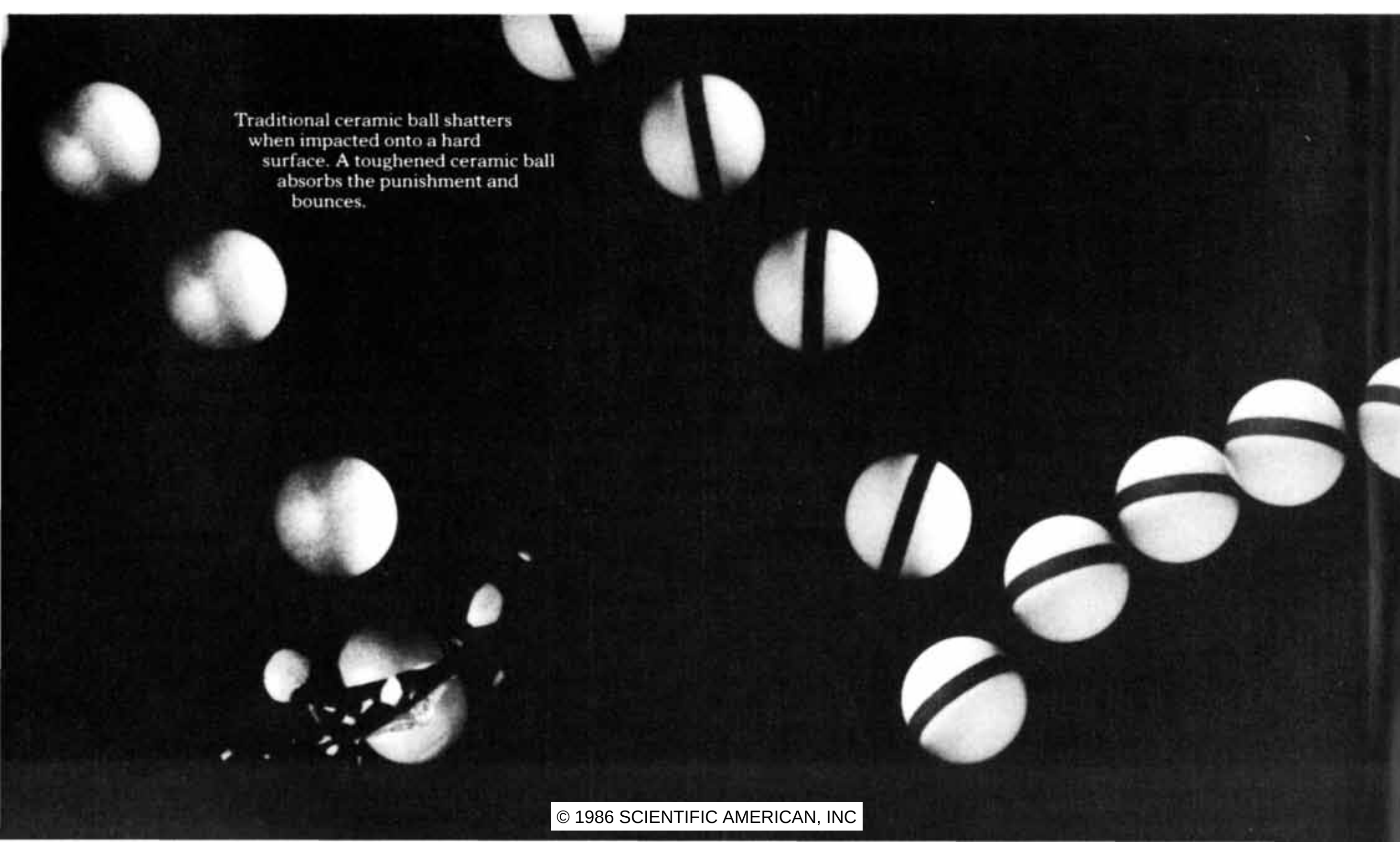
Toughened ceramics, on the other hand, fail more gracefully. A crack, once initiated, travels only a short distance before it is deflected or arrested by the toughening agent, using up a portion of the energy that produced the crack. This process is repeated many times, in effect allowing the ceramic to retain significant load-bearing capability after a crack is initiated.

In theory, there are several ways to toughen ceramics. At present, GTE is studying two of the most promising:

Tougher by a whisker.

Dispersing discrete solids throughout the ceramic matrix increases fracture toughness by promoting crack deflection. We are currently studying the effect of dispersoid size and shape (equiaxed particles vs. whiskers), as well as composition.

For example, SiC and TiC whiskers are being studied as dispersoids in silicon nitride, alumina and other ceramic matrices with great success. Toughness has been improved by as much as 40%.



Traditional ceramic ball shatters when impacted onto a hard surface. A toughened ceramic ball absorbs the punishment and bounces.

toughened ceramics.

Second Phase toughening.

Utilizing a mechanism similar to precipitation hardening in metals, ceramics can be toughened by the precipitation of a second phase. Optically transparent yttria radomes are being toughened with lanthana precipitates by as much as 35%. We are also studying transformation toughening dispersants in silicon nitride.

The golden age of ceramics.

GTE scientists are developing ceramics that are more and more "forgiving," more resistant to stress-induced cracking. Ceramics that can "carry the load" after a crack occurs.

There are already on the market tough ceramic cutting tools that outperform the best carbide tools. And now, through GTE research, toughened ceramic radomes, wear parts and heat-engine components may soon be commonplace.

To learn more about GTE's activities in this field, you are invited to request any of the papers listed at right. Write GTE Marketing Services Center, Department TC, 70 Empire Drive, West Seneca, NY 14224, or call 800-833-4000.

GTE

Pertinent Papers

W.H. Rhodes, J.G. Baldoni, and G.C. Wei, "The Mechanical Properties of La_2O_3-Doped Y_2O_3," presented at the Annual Meeting of The American Ceramic Society, May 8, 1985, Cincinnati, OH (Paper 175-B-85).

A.E. Pasto, "Toughened Ceramics: An Industry Viewpoint," outline of presentation at the topical symposium "Recent Developments in Ceramic Science," Sandia National Laboratories, October 1985.

S.T. Buljan, J.G. Baldoni, and M.L. Huckabee, "Si_3N_4-SiC Composites," to be published in The American Ceramic Society Bulletin, *1987.*

J.G. Baldoni, S.T. Buljan, and V.K. Sarin, "Particulate Titanium Carbide-Ceramic Matrix Composites," 427-58, Inst. Phys. Conf. Ser. #75: Clap Adam Hilger Ltd., Bristol, England.

S.T. Buljan, J.T. Neil, A.E. Pasto, J.T. Smith and G. Zilberstein, "Silicon Nitride Ceramics and Composites: A View of Reliability Enhancement," presented at the International Conference on Non-Oxide Technical and Engineering Ceramics, Limerick, Ireland, July 10-12, 1985. Conference proceedings to be published by Elsevier Applied Science, 1986.

The Functional Architecture of the Retina

Dozens of kinds of cells have specialized roles in encoding the visual world. New techniques have made it possible to study the arrangement and interconnections of entire populations of cells

by Richard H. Masland

The retina encodes the visual world. It transforms optical images into trains of nerve impulses, which are then conducted along the optic nerve to the brain. The brain interprets those signals to generate visual perception: a subjective sense of the shapes, colors and movements that surround the observer. The retina is more than just a bank of photocells, however. This thin sheet of neural tissue at the back of the eye is an outpost of the central nervous system. Its circuits of interconnected neurons, or nerve cells, carry out a form of image analysis: certain features of the raw visual input are accentuated and other features are downplayed.

The effort to understand how light signals are transduced into neural activity and how that activity is transformed has intrigued neurobiologists for many years. The general nature of the retina's encoding of the visual world—the relation of its output to its input—was established first. A skeletal view of the means by which the coding is accomplished was attained by the early 1970's. The retina was known to be composed of five main classes of neurons. They are connected to one another by synapses: points of close apposition where the chemical messengers called neurotransmitters are released by one neuron to affect the next neuron. Three of the five classes of retinal neurons form a direct pathway from the retina to the brain. These are the photoreceptors (the rod cells and cone cells), the bipolar cells and the ganglion cells. The remaining two classes of retinal neurons, the horizontal and the amacrine cells, form laterally directed pathways that modify and control the message being passed along the direct pathway.

Until recently it seemed that these five classes of cells exactly defined the retina's functional elements. Each class of cells was thought to carry out a single kind of task; to understand the retina's internal codes, neurobiologists should have only to decipher the interactions of these relatively few basic elements. Now, instead, it has become clear that the retina has many more than five functional elements. The five cell classes harbor subtypes so distinctive that the true number of functional elements may be as high as 50.

The task, then, is to learn why so many cell types are required. It is very far from having been completed, but in the past few years important progress has been made. A significant advance has been the development of ways to examine the shapes and the arrangement in the retina of entire populations of nerve cells.

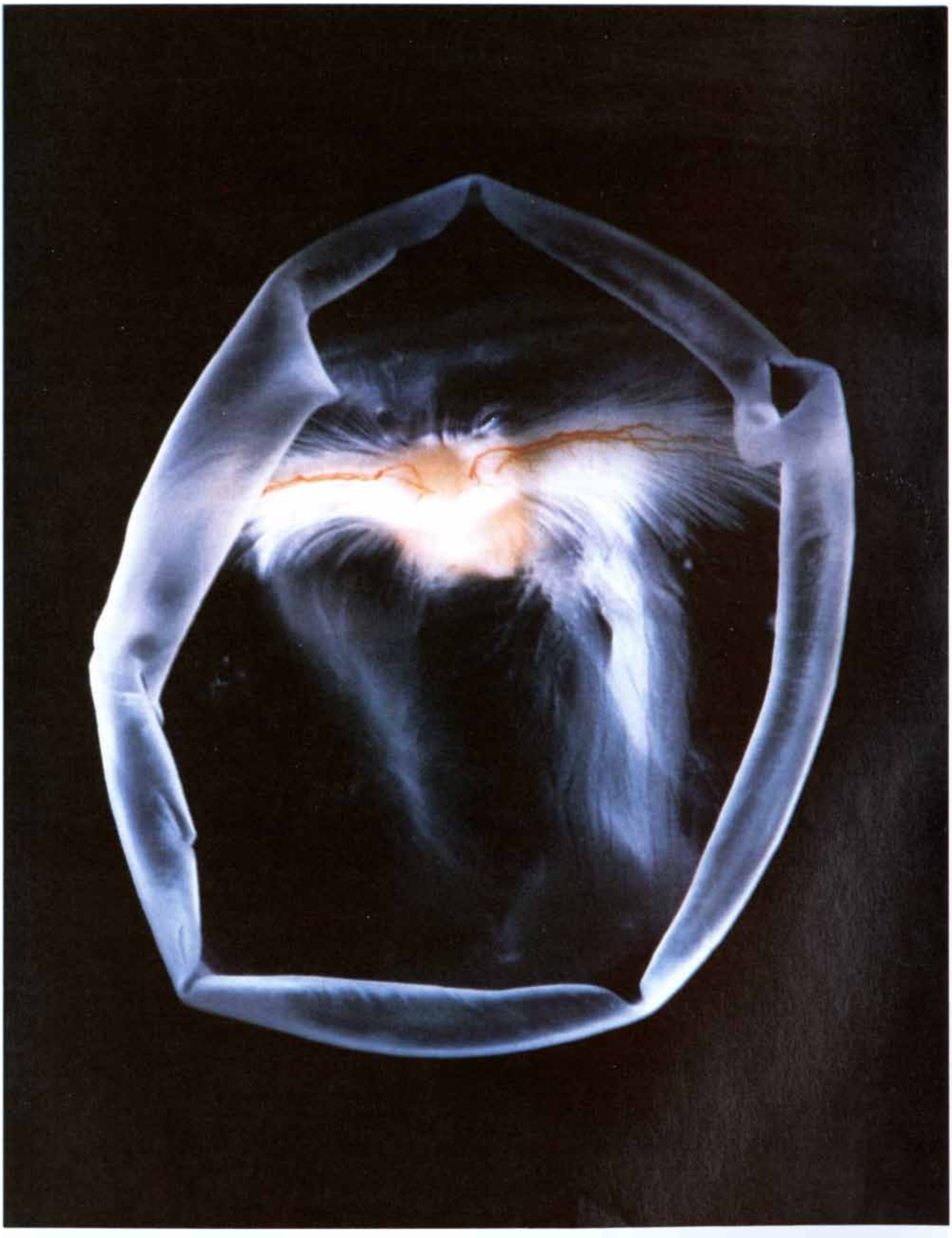
An understanding of precisely how the mammalian retina codes visual information began in 1952 with a series of experiments on ganglion cells, the cells whose axons form the optic nerve. These first experiments, carried out by Stephen W. Kuffler at the Johns Hopkins University School of Medicine, sought to answer the question: How does the electrical activity of ganglion cells change in response to light? Kuffler recorded the electrical activity of single ganglion cells. He found that most of them fire a continuous stream of action potentials, or nerve impulses, along their axons even in the absence of light. A ganglion cell responds to the presence of light by markedly increasing or decreasing its rate of firing.

Mapping the surface of the retina of anesthetized cats with small spots of light, Kuffler stimulated specific regions of the retina and measured the electrical response of individual cells. He found that each ganglion cell has a precise area of the visual field to which it responds: its receptive field. The size of the receptive field varies, but it is generally quite small. In the cat the centers of the smallest receptive fields occupy a region of the retina about 120 micrometers (thousandths of a millimeter) in diameter. If the cat looks at a wall two meters away, each such ganglion cell reports on an area roughly a centimeter in diameter.

The cells Kuffler studied can be divided into two populations: those that are stimulated and those that are inhibited when the center of their receptive field is stimulated. Moreover, the centers act in opposition to the region encircling them: the surround. If stimulation of the receptive field's center excites the cell, stimulation of the surround inhibits it. Conversely, if the center is inhibited by light, the surround is excited by it. In other words, the surround is "antagonistic"; the cell carries out a simple form of contrast enhancement.

A very different kind of ganglion cell was soon identified: the directionally selective cell, whose receptive field responds to the direction of a moving stimulus. (It turned out to be only one of a series of cells having complex stimulus selectivities, but it is the most abundant and well studied of them and can serve as a prototype.) Directionally selective cells were discovered in the frog by Jerome Y. Lettvin, Humberto R. Maturana, Walter H. Pitts and Warren S. McCulloch of the Massachusetts Institute of Technology. It was not until several years later, however, that the exact properties of these cells were elucidated in a series of classical experiments carried out on rabbits by Horace B. Barlow and William R. Levick of the University of Cambridge.

The behavior of a directionally se-



ISOLATED RABBIT RETINA, a sheet of neural tissue lining the inside of the back of the eye, is seen in a photograph made by the author. The retina, whose average thickness is only about a tenth of a millimeter, is essentially transparent. Light passes through it to strike the rod and cone cells, which are near its back surface. These photoreceptors transduce the light signal into a neural signal, which is relayed by successive layers of retinal neurons; ultimately the signal excites ganglion cells, whose axons form fibers (*white filaments*) that join to form the optic nerve. Two pairs of blood vessels are seen crossing the front surface of the retina. Methods for isolating the living retina from the eye were first developed by Adelbert Ames III of Massachusetts General Hospital.

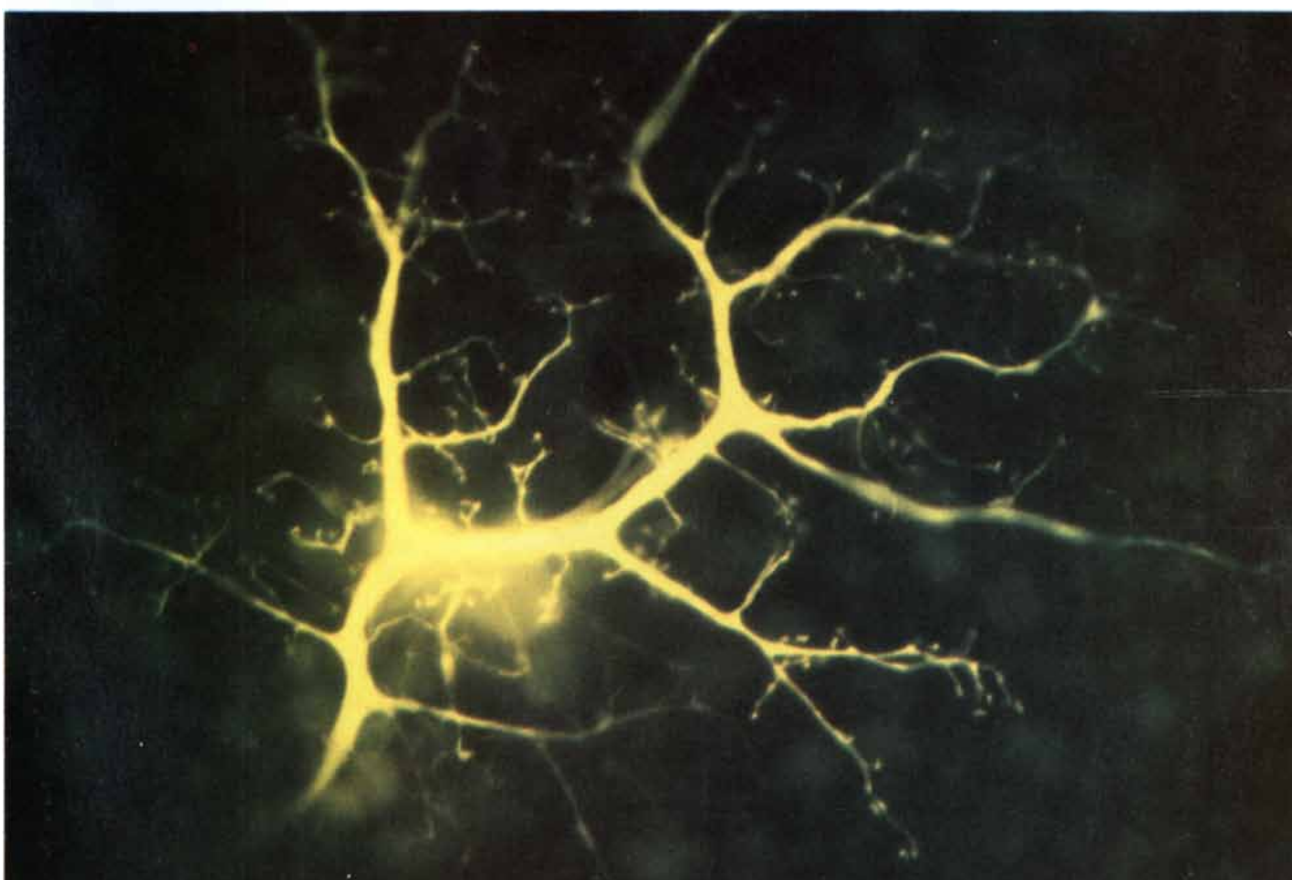
lective cell varies depending on whether the light stimulus is moving or stationary. If one maps the cell's receptive field with a stationary light, the cell responds uniformly throughout its receptive field. When a spot of light is moved slowly across the cell's receptive field, however, the behavior of the cell changes dramatically: it fires a sustained train of action potentials when the spot moves in one direction and fires little or not at all when the spot moves in the opposite direction. These early results defined basic questions about retinal circuits. What establishes a ganglion cell's receptive field? How can neuronal machinery transform the optical input so selectively? To find answers one must look at the retina's internal components in more detail.

As I mentioned above, the retina can be said to have two sets of neurons: those that establish a direct pathway from the source of light to the optic nerve and those that make lateral connections. It was immediately obvious that the antagonistic surround observed for some ganglion cells could be accounted for if the retina's laterally conducting neurons function in opposition to the through-pathway neurons. In other words, when the center of a ganglion cell responds to light (by way of the direct pathway), its surround (which acts in opposition) might be driven by stimuli transmitted by way of horizontal or amacrine cells. A role for horizontal cells received crucial support when it was learned that bipolar cells have antagonistic surrounds. Since the bipolar cell is the only cell that conducts from the outer retina (where horizontal cells reside) to the inner retina, no other machinery was needed to explain the existence of antagonistic surrounds.

Other lateral neurons, the amacrine cells, regulate the behavior of certain ganglion cells (including the directionally selective ones) that exhibit a transient response to stimulation by light. When light strikes the retina, these cells immediately fire a burst of action potentials, but they cease firing in the presence of continued light stimulation. Transient behavior is not exhibited to this degree by photoreceptor, bipolar or horizontal cells. It is, however, characteristic of many of the amacrine cells. One role of amacrine cells, then, is presumably to sharpen the transient responses of certain ganglion cells, including the directionally selective ones.

Amacrine cells must have more roles than that, however. These cells come in a bewildering variety; a single retina may contain as many as 30 morphologically distinct types. If amacrine cells had no function other than to make the response of some ganglion cells transient, one might reasonably expect there to be only one kind of amacrine cell, or certainly no more than a few kinds.

The diversity of amacrine cell shapes had been reported as far back as 1892 by the great Spanish neuroanatomist Santiago Ramón y Cajal. For a long time his findings were disregarded. Among other possibilities, one could always argue that the different cell shapes he described were merely variants: cells that look different but have the same function. It was not until the late 1960's, when Berndt Ehinger of the University of Lund in Sweden began studying amacrine cell biochemistry, that the true diversity of amacrine cells was recognized. Ehinger applied to the retina newly devel-



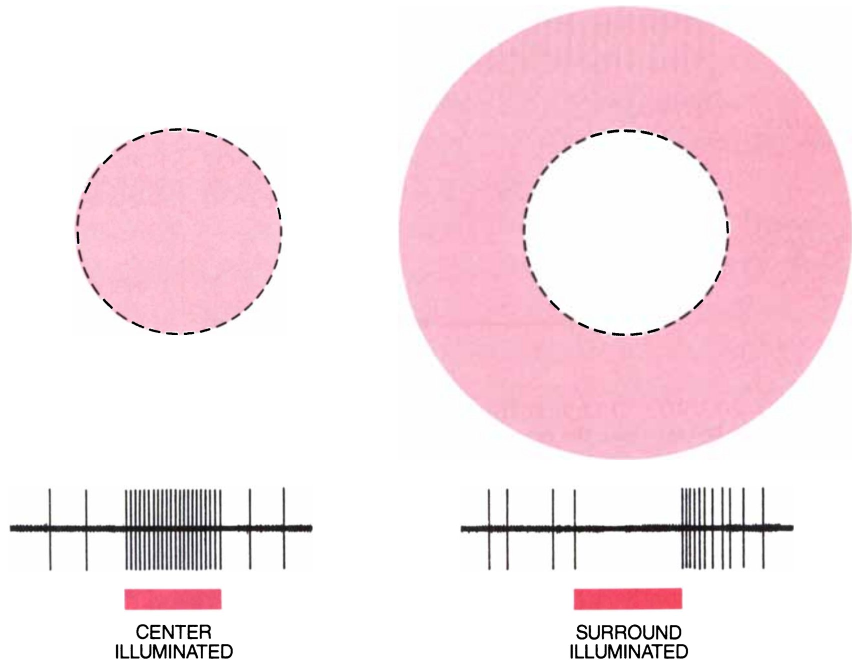
RETINAL NEURON, a horizontal cell, was injected with the fluorescent dye Lucifer Yellow, which stains all the cell's processes. A terminal arborization of the cell is seen in a photomicrograph made by Julie H. Sandell. Information flows laterally along such a cell's branches and small twigs to the buttonlike terminal boutons, which make synaptic contact with other cells in the retina.

oped methods for identifying neurotransmitters in thin sections of neural tissue. He found that a large number of the neurotransmitters known to be present in the brain were also present in retinal neurons. The surprising thing was that all these neurotransmitters were found in one or another amacrine cell, each neurotransmitter in a subset of the amacrine cells. This implied that there are many distinct subsets of amacrine cells.

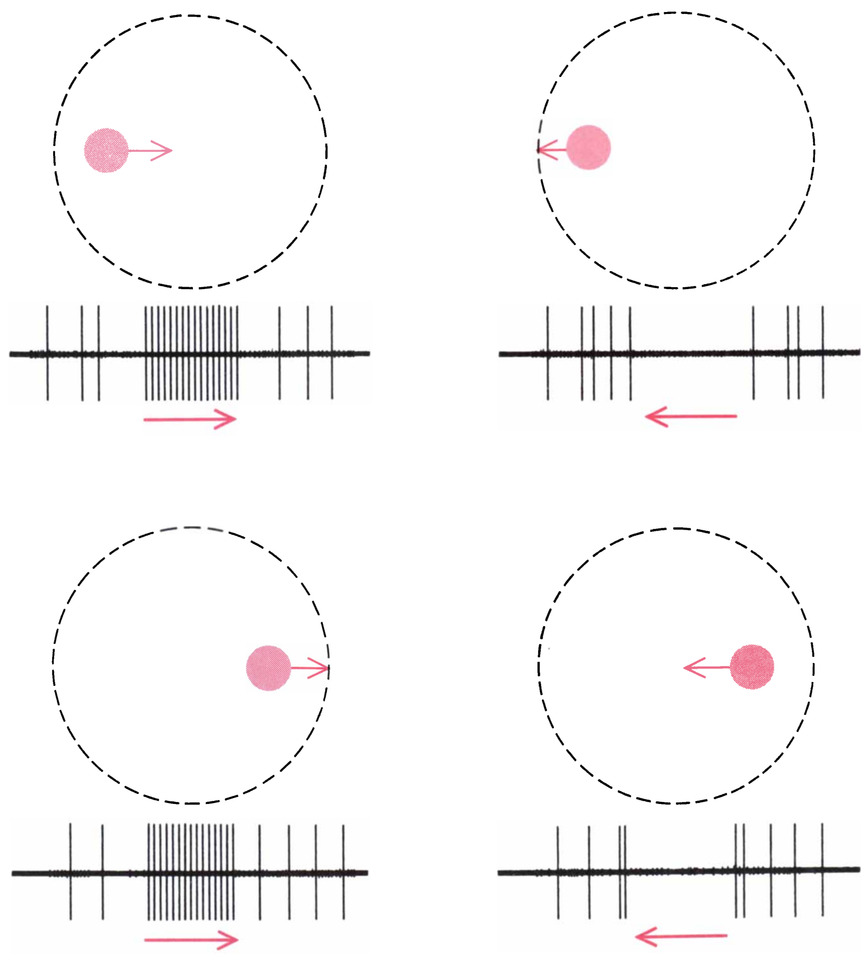
At first Ehinger's results were viewed with some skepticism. The methods were new, and one could argue that they had overestimated or underestimated the amacrine cell populations. Yet subsequent studies by other investigators led to the same results. Among these was a study, carried out by John W. Mills and me at Massachusetts General Hospital, of amacrine cells that synthesize the neurotransmitter acetylcholine. We incubated isolated rabbit retinas in the presence of radioactively labeled choline, the precursor from which the cells synthesize acetylcholine. Our next objective was to immobilize acetylcholine for autoradiography (a method for localizing radioactivity in tissue sections) so that we could determine which cells had synthesized acetylcholine. To do this we used a fast-freezing method. Samples of living retinas were plunged directly into propane at −180 degrees Celsius, the tissues were freeze-dried and then prepared for autoradiography by methods that keep them from coming into contact with moisture.

These methods are tedious, but they have a special advantage: the tissue's acetylcholine is immobilized within a few milliseconds by the freezing. When it is subsequently located by autoradiography, it must still be within the cell where it originally belonged. The technique thus eliminates some of the uncertainties that attended earlier methods for identifying neurotransmitters in cells. The labeling was sharp and distinctive: a cell was either densely labeled or not labeled at all. Acetylcholine was present only in a very small subset of amacrine cells. As time went on, other laboratories also confirmed Ehinger's fundamental finding. As he had suggested, different neurotransmitters are confined to small subsets of amacrine cells.

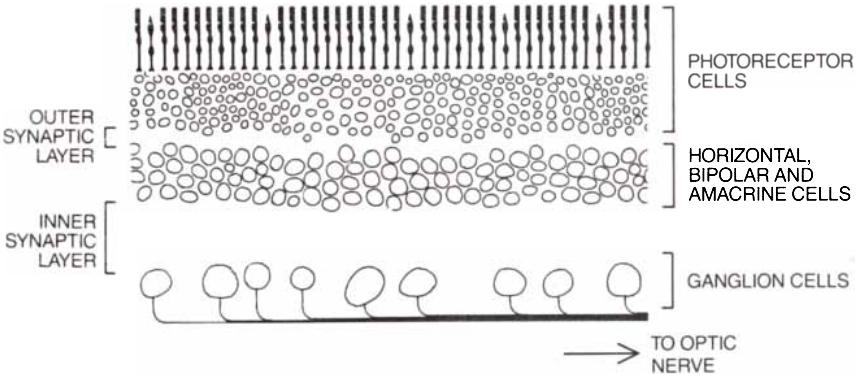
Knowing this, one could try to link particular neurotransmitters to morphologically distinct types of amacrine cells, such as those described by Ramón y Cajal. It was quickly learned that amacrine cells having differently shaped dendritic trees could be matched with particular neurotransmitters. Among the evidence, more-



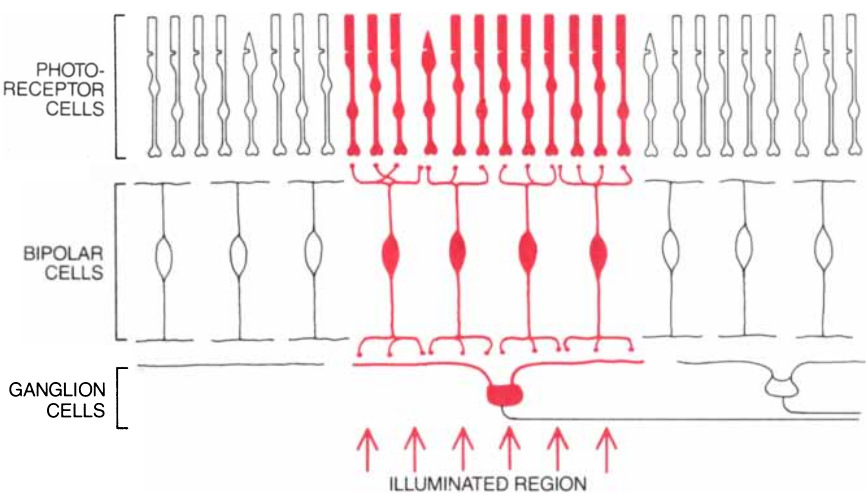
CONCENTRIC RECEPTIVE FIELD has a central region in which stimulation by light has one effect on a ganglion cell's firing (recorded by a microelectrode measuring the activity of a single axon in the optic nerve) and a surrounding region in which the effect is the opposite (*top*). In the case of an on-center cell, stimulation of the center causes an increase in activity (*bottom left*), stimulation of the surround a decrease (*bottom right*).



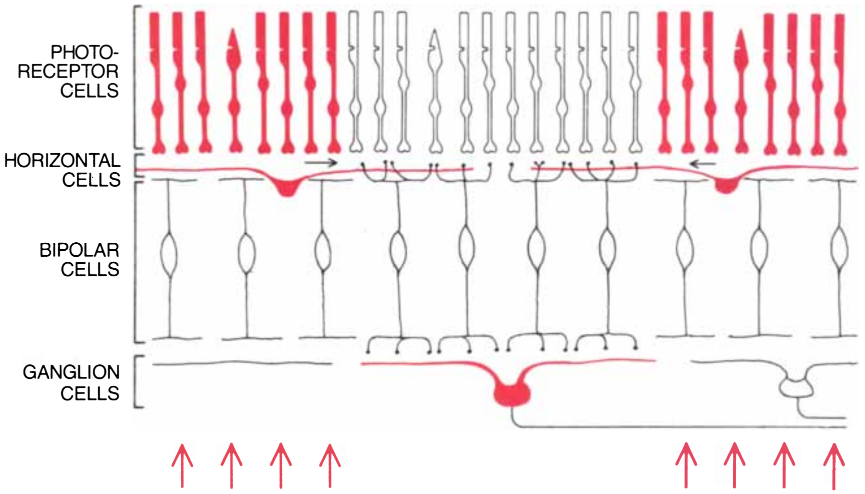
DIRECTIONALLY SELECTIVE CELL responds when a spot moves in one direction (to the right in this diagram) and is inhibited when the spot moves in the opposite direction. The direction of movement is detected wherever the spot may fall in the receptive field.



RETINAL-NEURON CELL BODIES are arranged in three layers. In addition there are two synaptic layers, where the processes of the cells (axons and dendrites), which are not seen in this schematic diagram, intertwine and make synaptic contact with one another.



RETINA'S PRIMARY THROUGH PATHWAY is provided by three kinds of neurons: photoreceptor cells, bipolar cells and ganglion cells, which are shown in this expanded view. A retina composed of these cells could transmit information about light, but its ganglion cells would display no selective features; they would have only a "center" response.



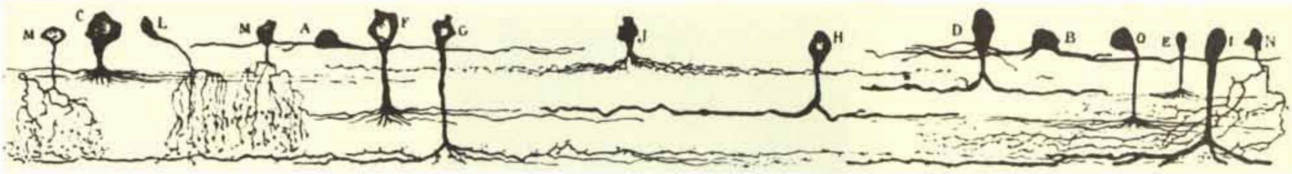
LATERAL INTERACTIONS can account for such selective features as a receptive field's antagonistic surround. Suppose the horizontal cells have an effect (*black arrows*) on the bipolar cells that is opposite to the direct effect of the photoreceptor cells. Then the final signals transmitted to a ganglion cell will have a center and an antagonistic surround. The actual horizontal-cell connections that mediate this effect are still incompletely known.

over, was work by Nicholas C. Brecha and Harvey J. Karten of the State University of New York at Stony Brook showing that not only the traditional neurotransmitters but also many of the body's neural peptides (a large family of newly discovered peptide neurotransmitters) are present in distinct amacrine cells. This finding increased the diversity of amacrine cells beyond all expectations.

The discovery that shape and synaptic chemistry are interrelated in amacrine cells had a second and crucially important consequence. It removed any remaining doubt that amacrine cells having different shapes have different biological functions. When a neurotransmitter binds to a receptor on a postsynaptic cell, that cell undergoes specific biochemical and physiological changes. If various amacrine cells have both different shapes (which imply different connections within the retina) and different neurotransmitters, they must have differing biological roles. What role could 30 different amacrine cells possibly have? Each type undoubtedly has its specific function, but there seem to be far more kinds of cells than there are jobs for them to do. In my laboratory and in others some of these jobs are now beginning to be revealed. Here I shall describe work with four very different amacrine cells.

The first of these cells has already been mentioned. It is the cholinergic, or acetylcholine-containing, amacrine cell that was identified by autoradiography. One drawback to autoradiography was that it reveals the cell body of a neuron but not the processes: the axons and dendrites that transmit and receive messages. In fact, the major anatomical method for seeing the processes was the Golgi technique—the very method Ramón y Cajal used in the 19th century. That method cannot be controlled: it stains an entire cell, but the investigator cannot choose the cell ahead of time.

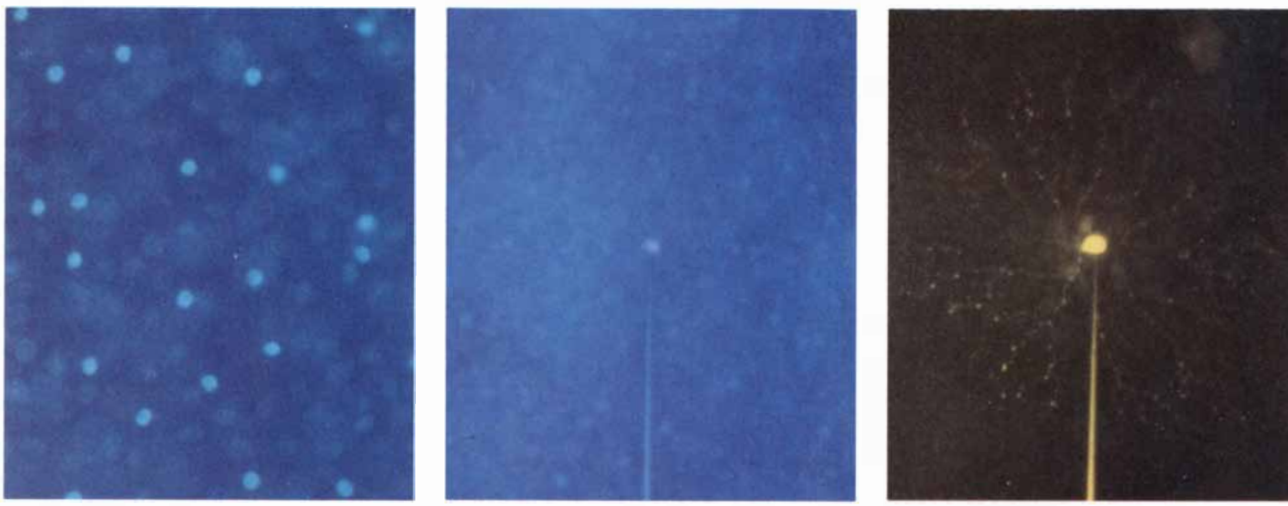
A substitute for the Golgi technique was suggested by my discovery that a fluorescent molecule, 4,6-diamidino-2-phenylindole (DAPI) accumulates selectively in the cell bodies of the cholinergic neurons. If one treats an entire retina with DAPI and then mounts the retina flat, the cholinergic neurons fluoresce beautifully against the darker background. This made it possible for Masaki Tauchi, a postdoctoral student, and me to guide a fine micropipette to a cell and inject it with a different fluorescent dye, Lucifer Yellow CH. The Lucifer Yellow diffuses through the cell's entire network of branches. The results were perfectly



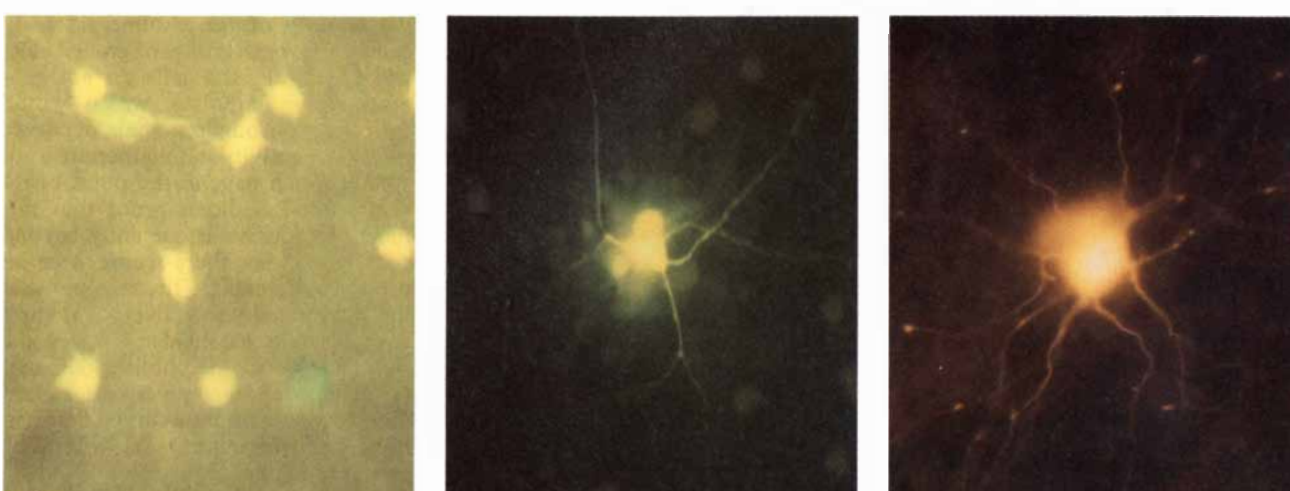
WIDE VARIETY of amacrine cells is displayed in this drawing, published in 1892 by the eminent Spanish neuroanatomist Santiago Ramón y Cajal, of amacrine cells in the carp retina. He was able to distinguish 14 cell types on the basis of their shape alone.

consistent: all the fluorescent cells had the same shape and branching pattern.

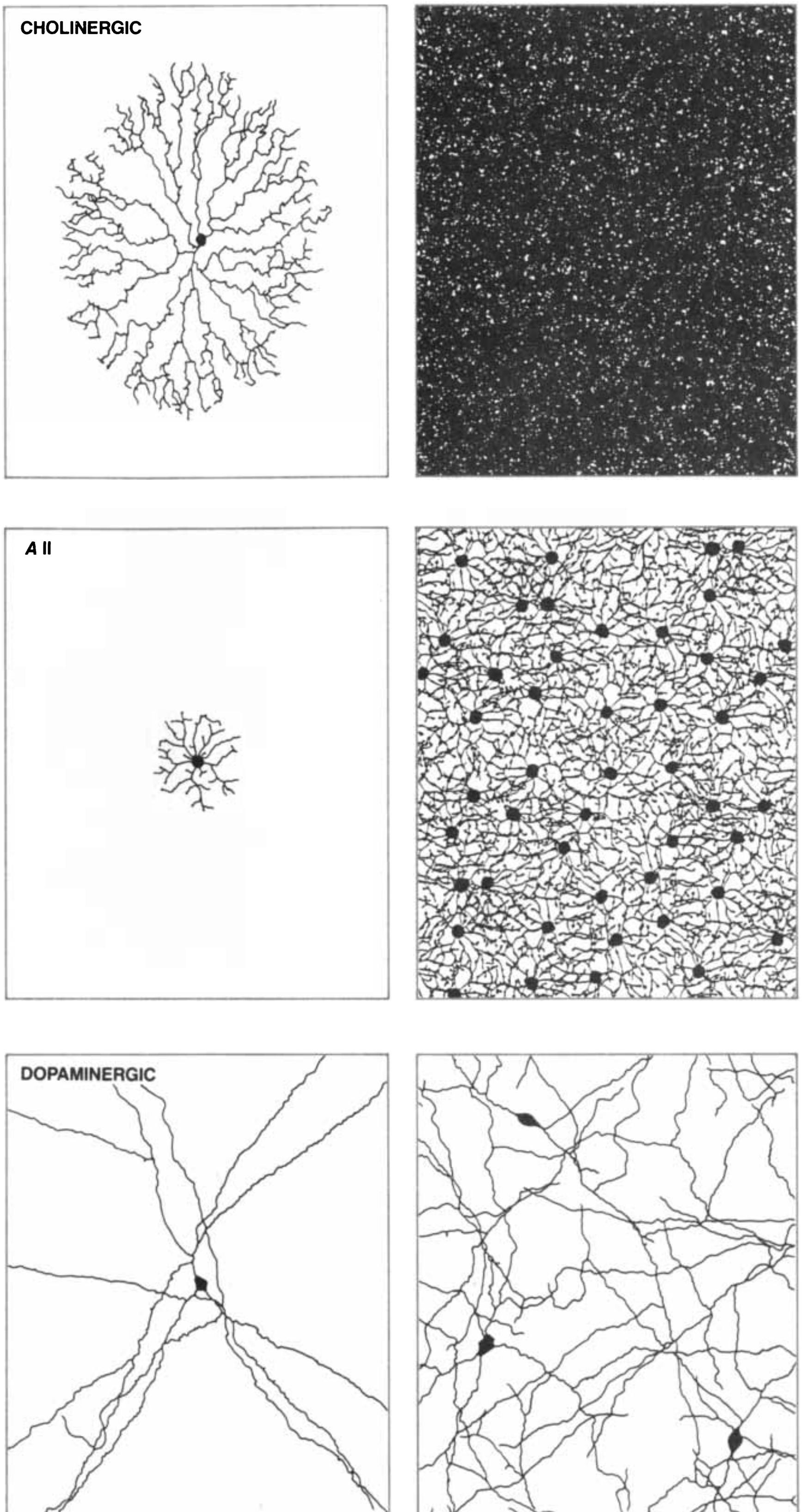
This method not only showed us the true shape of the cholinergic cells but also enabled us to reconstruct the mosaic they form as they blanket the retina. To see why this was important one must again think about the job the retina does. The retina must, above all, preserve spatial resolution; any failure would compromise the acuity with which the animal can see. The density and spatial arrangement of the retina's intermediate elements—through which visual information must pass—thus provide important information about the role those elements can play within the system as a whole. Pioneering analyses of retinal cell mosaics had been done by Brian B. Boycott and Heinz Wässle at King's College London for



CHOLINERGIC AMACRINE CELLS (which contain the neurotransmitter acetylcholine) are stained by a technique devised by Masaki Tauchi and the author. A retina is mounted flat as an intact sheet. The cholinergic amacrine cell bodies selectively accumulate DAPI, a blue fluorescent dye (*left*). This makes it possible to guide a micropipette that has been filled with another dye, Lucifer Yellow CH, into an individual amacrine cell under illumination that makes both dyes fluoresce (*center*). Under different illumination (*right*) only the Lucifer Yellow fluoresces, revealing the distinctive branching pattern of the cholinergic cell's dendrites.



TWO OTHER AMACRINE CELL TYPES are stained by the same method. A retina was treated in such a way that the cell bodies of dopaminergic amacrine cells fluoresce in green, those of indoleamine-accumulating amacrine cells in yellow (*left*). Injection of individual cells shows the specific dendritic shapes of dopaminergic cells (*center*) and indoleamine-accumulating cells (*right*).



MOSAICS of different amacrine cells cover the retina differently. Three cell types are shown, all in flat view. Individual cells are at the left; the drawings at the right show an approximation of the mosaic that would result if all the cells of a class were stained simultaneously. Cholinergic cells are numerous and their branching dendrites form an almost uninterrupted meshwork (*top*). The mosaic of AII amacrine cells is sparser (*middle*): there is more space between their dendrites. Dopaminergic cells are sparser still (*bottom*). The drawings were synthesized on the basis of average cell shape and density.

two cells that can be stained as entire populations by classical methods. Our fluorescence methods make it possible to study many other cells.

When Tauchi and I measured the size and density of the cholinergic cells, we found a surprising result. The cells overlap one another a great deal. In the peripheral retina, where the overlap is greatest, a point on the retinal surface is overlaid by the processes of some 140 cholinergic amacrine cells. Boycott and Wässle's work had suggested that the overlap would be slight. Such enormous redundancy was entirely unexpected.

A possible explanation for the redundancy was suggested by the kinds of stimuli the cholinergic amacrine cells can resolve. From previous electrophysiological work we knew that cholinergic amacrine cells excite certain ganglion cells, among them the directionally selective ones. We knew that these ganglion cells can detect the movement of very small spots—spots smaller than their dendritic field. The spots are also much smaller than the dendritic field of the cholinergic amacrine cells. Because most neurons simultaneously transmit the same message across all their synapses, an amacrine cell would be expected to release acetylcholine from all points on its dendritic tree whenever any one point was excited. How can a cell that is so spread out transmit precise information about stimuli smaller than itself?

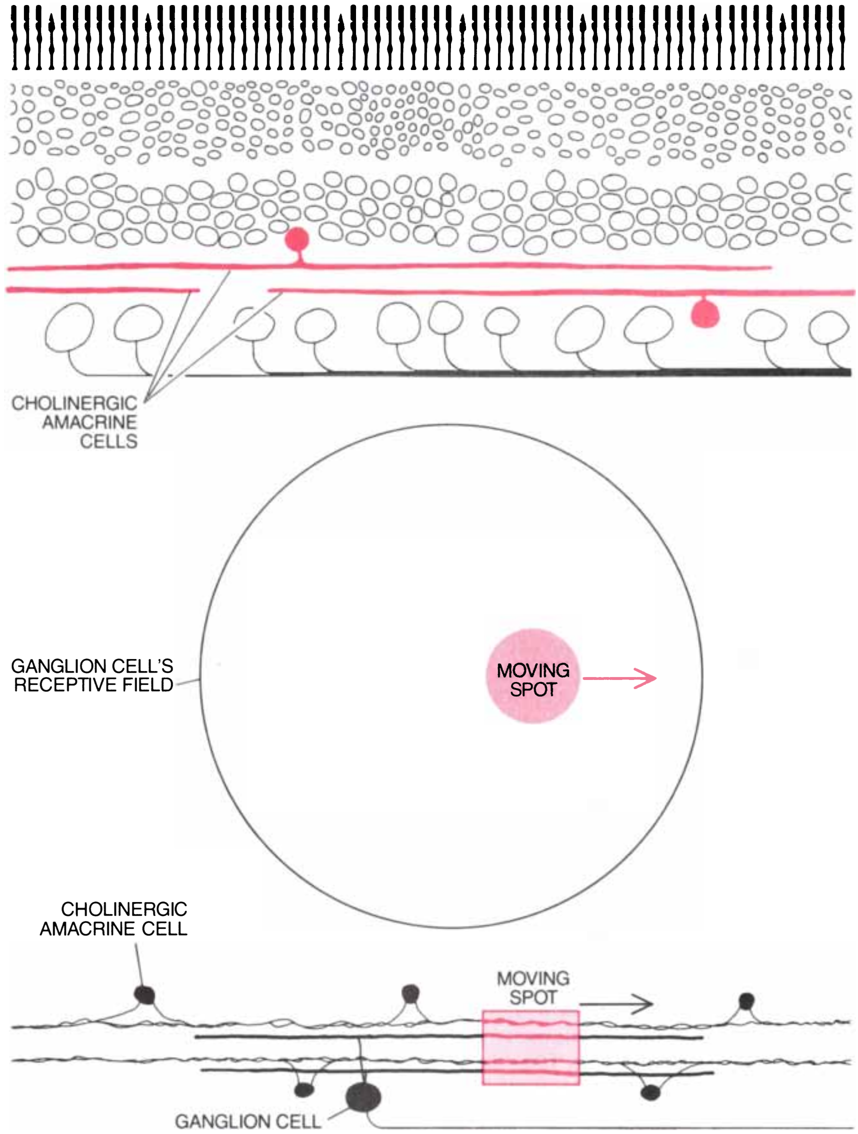
Our best guess is that the dendrites of the amacrine cell are electrically isolated, enabling one region to release acetylcholine to a ganglion cell locally without there being any release at more distant sites. This is consistent with the structure of amacrine cells, in which inputs and outputs exist side by side on the same process. The electrical activity of the cholinergic amacrine cells consists only of graded electric potentials; the cells do not generate nerve impulses, which would propagate throughout their dendritic trees. The activity should therefore not spread much beyond the point of input. Although direct proof that this mechanism exists is currently beyond our capabilities, there seems to be no other way to explain the resolution of the directionally selective cell. Moreover, such a mechanism makes the tight meshwork of cholinergic amacrine cell dendrites an actual advantage: a small stimulus can be detected no matter where it may fall within the ganglion cell's receptive field.

The second type of amacrine cell I shall describe is the AII cell, which differs from the cholinergic cell in that it has an extremely narrow lateral spread. David Vaney of the University of Cambridge was able to describe the mosaic formed by AII cells by adapting our fluorescence-guided injection method. These cells are abundant and cover the entire surface of the retina, but they are so small that their dendrites do not overlap much.

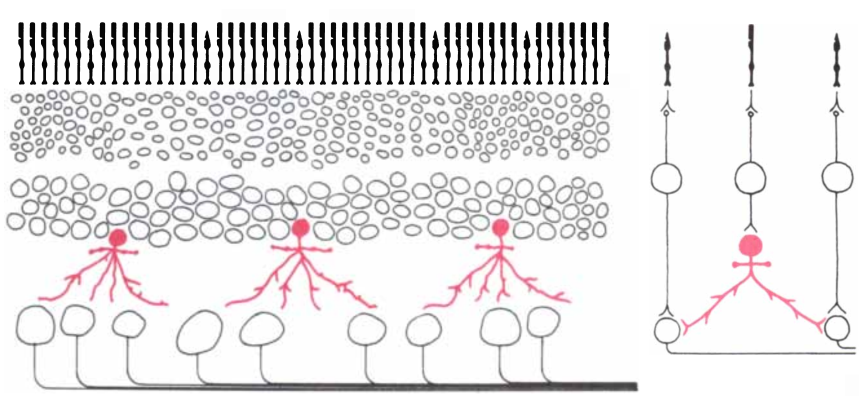
Several other workers have studied AII cells, including Helga Kolb and Ralph F. Nelson of the National Institutes of Health, Barbara A. McGuire, Peter Sterling and John K. Stevens of the University of Pennsylvania School of Medicine and Ramon F. Dacheux and Elio Raviola of the Harvard Medical School. As a result of their painstaking work there is now an answer to a question posed several years ago by Kolb and Edward V. Famiglietti: How can ganglion cells fire in response to dim light even though they do not receive a direct synaptic input from the bipolar cells that are activated by rod photoreceptor cells? Only rod cells respond to dim light, and cats (like most mammals) can see in dim light. Yet the anatomical evidence in cats indicated that their ganglion cells receive little or no direct input from rod-driven bipolar cells.

A major part of the answer lies with the AII amacrine cell. This cell has two apparent functions. Like many other amacrine cells, it transmits a transient response to light to the ganglion cells, thereby sharpening their response to the onset of stimulation. But it also serves to connect rod-activated bipolar cells to ganglion cells. By doing so it allows the ganglion cells to function under both bright and dim light conditions. In fact, the AII amacrine cell by virtue of these connections becomes part of the retina's through pathway. The flow of information is from rod photoreceptor to bipolar to AII amacrine to ganglion cell. It makes sense that the AII amacrine cells are small and densely packed: they thereby keep the level of acuity high along that crucial path.

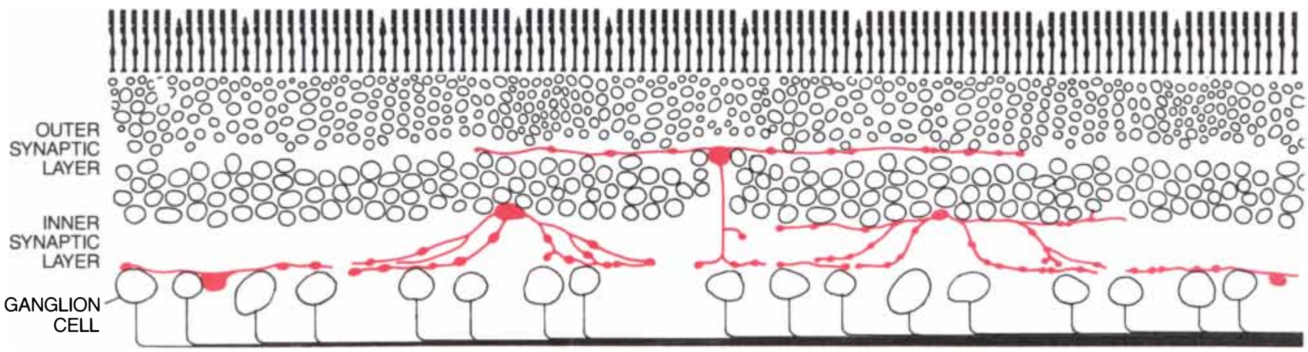
The third cell, like the first, is identified by its neurotransmitter, in this case dopamine. Tauchi and I found we could use the same kind of fluorescent injection method on these cells as on the cholinergic cells. What is notable about the dopaminergic cells is their scarcity in the retina. To give an example, the rabbit retina incorporates (in round numbers) some 350,000 ganglion cells and 300,000 cholinergic amacrine cells but only some 8,500 dopaminergic amacrine cells. Furthermore, the dendrites of the dopamine-containing cells are few and are thinly branched, resulting in a mosaic that is full of spaces, not densely packed as in



RESOLUTION of the cholinergic amacrine cell may be explained by the fact that its dendrites are both input and output processes and can act in isolation from the rest of the cell. Cholinergic amacrine cells are shown schematically (*color*) in a vertical section of the full retina (*top*). The edges of a moving spot in a ganglion cell's receptive field (*middle*) excite only particular regions of the mosaic of amacrine cell dendrites (*bottom*); the cholinergic dendrites thereupon excite the ganglion cell only in the illuminated region.



AII AMACRINE CELL (*color*) is positioned between rod-driven bipolar cells and ganglion cells (*left*). It provides a bridge between a rod-driven bipolar cell and a ganglion cell (*right*). In this drawing the cell's connections are shown schematically; an AII amacrine cell actually is connected to different types of ganglion cell by different pathways.



INDOLEAMINE-ACCUMULATING CELLS (*color*) **come in five shapes. Four are like other amacrine cells; one is unusual, bridging between the outer and inner synaptic layers. They all send processes to the lowest level of the inner synaptic layer, where most or all of them synapse with rod-driven bipolar cells. They appear to provide five pathways to a common end point.**

the case of the other amacrine cells discussed here.

Although the precise function of the dopaminergic cells is not known, the loose mosaic does clearly suggest that these cells are not involved in activities that require a high degree of spatial resolution. Whereas the exceptionally densely packed mosaic of the cholinergic amacrine cell makes possible the exact resolution of a small spot crossing the retina, in the dopaminergic cells a small spot would often fall on one of the holes in the mosaic. In other words, it is unlikely that the dopaminergic cell controls any of the obvious features of the ganglion cell's receptive fields. Electron microscope observations by John E. Dowling and Ehinger, working at Harvard, show that dopaminergic cells have synapses only with other amacrine cells. They have no direct connection to the retina's primary through pathway. Moreover, electrophysiological experiments indicate that the dopaminergic cells affect the activity of ganglion cells in a vague way that is hard to define. One possibility is that they have some rather diffuse function, such as mediating the overall excitability of the inner retinal neurons.

The last amacrine cell to be discussed is distinguished experimentally by its accumulation of chemical analogues of the neurotransmitter serotonin. Because serotonin and the analogues are indoleamines, the cells are termed "indoleamine-accumulating." After labeling the retina with a fluorescent serotonin analogue, 5,7-dihydroxytryptamine, Julie H. Sandell (a postdoctoral student) and I were able to examine these cells and study their morphology.

We found that this class of amacrine cells can be subdivided into five distinct morphological types. They have so many features in common, however, that it is best to consider them a family of cells rather than functionally independent cell types. There are many reasons for this. First, their dendrites share a family resemblance in size, shape and branching pattern that sets them apart from the dendrites of other amacrine cells. Second, they transport and accumulate the same serotonin analogue. Finally, and most important, their dendrites all participate in a dense plexus, or network, lining the innermost margin of the inner synaptic layer. There the cells make a characteristic synapse, called a reciprocal synapse, with a terminal of the rod-driven bipolar cell.

It is outside the plexus that the differences among these cells manifest themselves, and this is reflected in their synaptic connections and their overall shapes. Because the indoleamine-accumulating cells branch extensively and are in contact with many if not all rod-driven bipolar cells, they are thought to have a major influence on the pathway by which dim light passes through the retina. The logical conclusion is that the five indoleamine-accumulating cells represent five different pathways by which other retinal neurons can interact with rod bipolar cells.

As this small sample of the amacrine cells makes plain, the retina contains a great deal of extremely sophisticated circuitry. In fact, the complexity of the retina is dismaying even to neurobiologists. "The retina is supposed to be a simple system," colleagues typically complain. "How can it be more complicated than other brain structures?" To tease them one can respond that the retina seems complicated only because more is known about it. (The unspoken fear, of course, is that the retina is not notably complicated; if a structure thought to be a prototype of simplicity is really this intricate, what must the cerebral cortex be like?) There are more serious answers, however.

The first answer is that the overall task of the retina, the conversion of light signals into meaningful trains of nerve impulses, demands complex machinery. This article has addressed only one aspect of the retina's coding of visual input: the organization of two kinds of ganglion cell receptive fields. I have completely omitted, for example, the retina's remarkable ability to alter its own sensitivity levels. It does so through internal adjustments in responsiveness, which take place at every stage in the retinal hierarchy. It is these adjustments that enable most mammals to see effectively in both sunlight and starlight, a 10-billionfold range in light intensity. Retinas are also able to transmit information to the brain about the wavelengths of the light they receive, enabling human beings and some other vertebrates to see colors. Given the sophisticated tasks accomplished by the retina, it makes sense that its wiring is complicated.

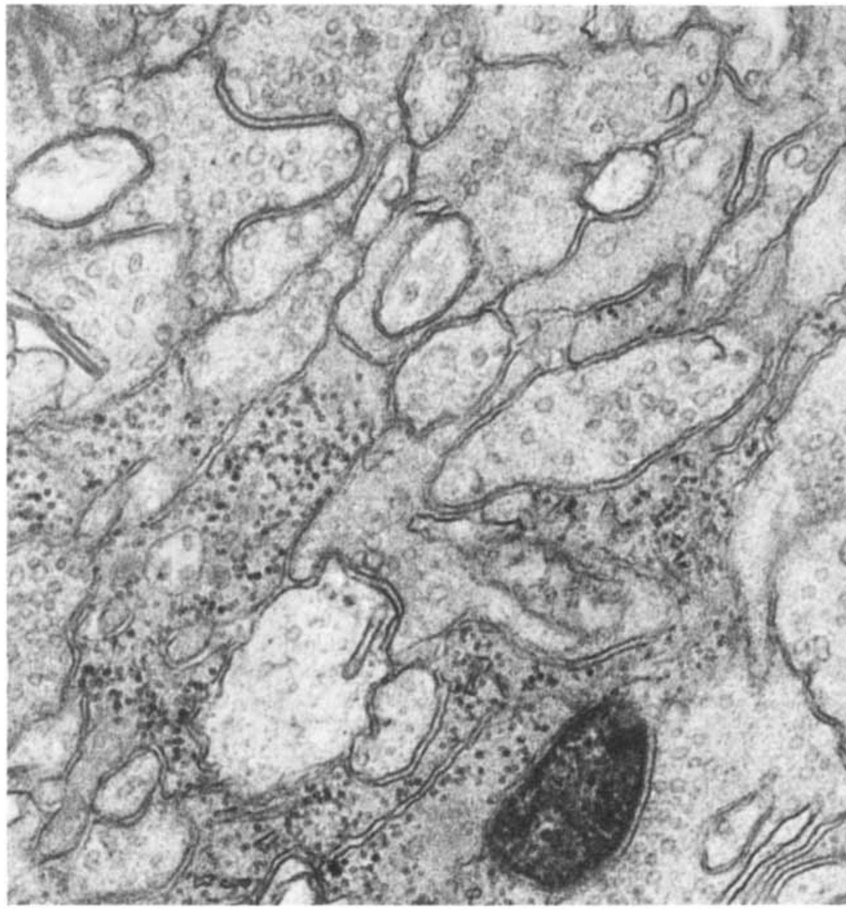
A second answer is that sophisticated circuitry may be required even for what seem to be straightforward tasks. There was no way to predict, for example, that the rod bipolar cells would be found not to synapse directly with ganglion cells. Making the connection by way of the AII amacrine cells seems illogical. Not only is it an indirect mode of communication but also it adds lateral spread (admittedly a small one) to the visual pathway and thereby threatens visual acuity. If cone bipolar cells make contact with ganglion cells directly, why should rod cells be any different? The only conclusion to be drawn is that the presence of AII amacrine cells solves some problem we are not even aware of yet.

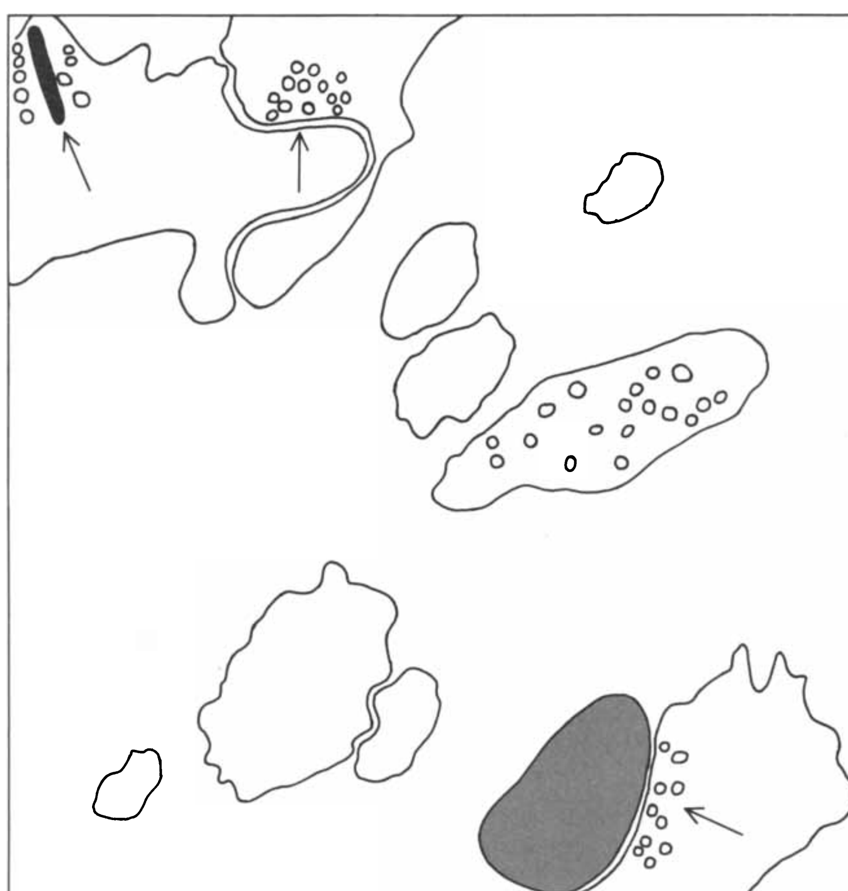
The third and perhaps the best answer is that the retina packages a lot of complexity within a small area. All parts of the central nervous system operate under constraints of space, but rarely is space as limited as it is in the retina. There are two reasons for this:

retinal cells must be packed together very tightly to maximize visual acuity, and the retina must be very thin (in the rabbit about a tenth of a millimeter) so that light can penetrate to the rods and cones. One square millimeter of the rabbit's central retina overlies 9.4 linear meters of cholinergic cell dendrites alone. Add to this the dendrites of all the other neurons in the retina, and the density of cell processes is staggering. It is not surprising, then, that retinal neurons are smaller than most other neurons. If their cell bodies were the same size as those of the human brain, the human eye would be almost the size of a tangerine. In short, the retina is a triumph of miniaturization.

Research on the architecture of the retina is about to enter a new and promising phase: the reconstruction in three dimensions of functionally related groups of neurons. The methods we have developed for revealing the shapes of retinal neurons can readily be combined with electron microscopy. They simplify the electron microscopist's task by providing a guide for following individual dendrites through the thickets of each synaptic layer. Furthermore, the same miniaturization that usually presents a barrier to the investigation of retinal cells becomes an advantage when one attempts three-dimensional reconstruction. The close packing of retinal cells makes it easier to examine entire retinal circuits in electron micrographs of serial tissue sections and then to reconstruct the circuitry.

Although this kind of investigation is now in its infancy, the value of space-filling models is clear. Such neural reconstructions are the ultimate description of the architecture and interconnections of retinal cells. Other kinds of inquiry will be necessary before one can say how those structures function electrically and chemically. When the structures are known, however, the work begun by Ramón y Cajal nearly a century ago will have been completed.





INNER SYNAPTIC LAYER is enlarged some 50,000 diameters in an electron micrograph (*top*) made by Ramon F. Dacheux and Elio Raviola of the Harvard Medical School. The dendrites of a number of neurons can be seen; a few of them are indicated on the map (*bottom*). The dark material in one dendrite identifies the dendrite as that of an indoleamine-accumulating cell. The usual fluorescent stain was converted into an electron-dense material by a technique developed by Sandell and the author. Vesicles (*small circles*) containing neurotransmitters are seen at synapses (*arrows*).

How Roots Respond to Gravity

A century ago botanists discovered that a root turned on its side curves downward rapidly in response to gravity, but the mechanism underlying this behavior is only now beginning to be understood

by Michael L. Evans, Randy Moore and Karl-Heinz Hasenstein

In gravity-free outer space there is no up or down. When a plant is moved from some arbitrarily designated vertical position to a horizontal one, the stem and the root continue to grow straight; they do not return to their initial positions. On the earth, plants behave very differently. They exhibit gravitropism (which used to be called geotropism): their parts reorient in response to being "gravistimulated," or repositioned with respect to gravity. After a vertically oriented plant is turned on its side the shoot curves upward, away from the direction of gravity's pull. The root, which is the subject of our investigations, curves downward, in the direction of gravity. In many instances a root begins to reorient within 10 to 30 minutes of being gravistimulated.

Although a casual observer might suspect that such downward curvature by roots is merely a search for water or a passive response to gravity, neither is the case. Roots of some plants do seek water, but gravity has an independent effect on the direction of root growth. Moreover, growing roots, which are strong enough to penetrate packed soil, are certainly strong enough to resist the direct pull of gravity; root gravitropism arises from some other mechanism.

Botanists have attempted to uncover this mechanism since before the time of Charles Darwin, whose 1881 book *The Power of Movement in Plants* included detailed descriptions of gravitropism. It is nonetheless only with the advent of the U.S. space program, and a consequent increase in funding for research into the effects of gravity and zero gravity, that investigators have made significant progress in describing the processes that underlie gravitropism in roots.

Gravitropism can be divided into three phases: perception, transduction and response. In roots the perception, or initial detection, of gravity appears to occur in the cap, which is the terminal half millimeter of the root. The response—an altered growth pattern leading to downward curvature—takes place somewhat behind the cap, in the "elongating zone," where, as the name implies, cells elongate. Transduction, the intermediate (and most mysterious) phase, almost certainly involves some kind of communication between the cap and the elongating zone, but the exact nature of the interaction between the two regions is only now beginning to be deciphered.

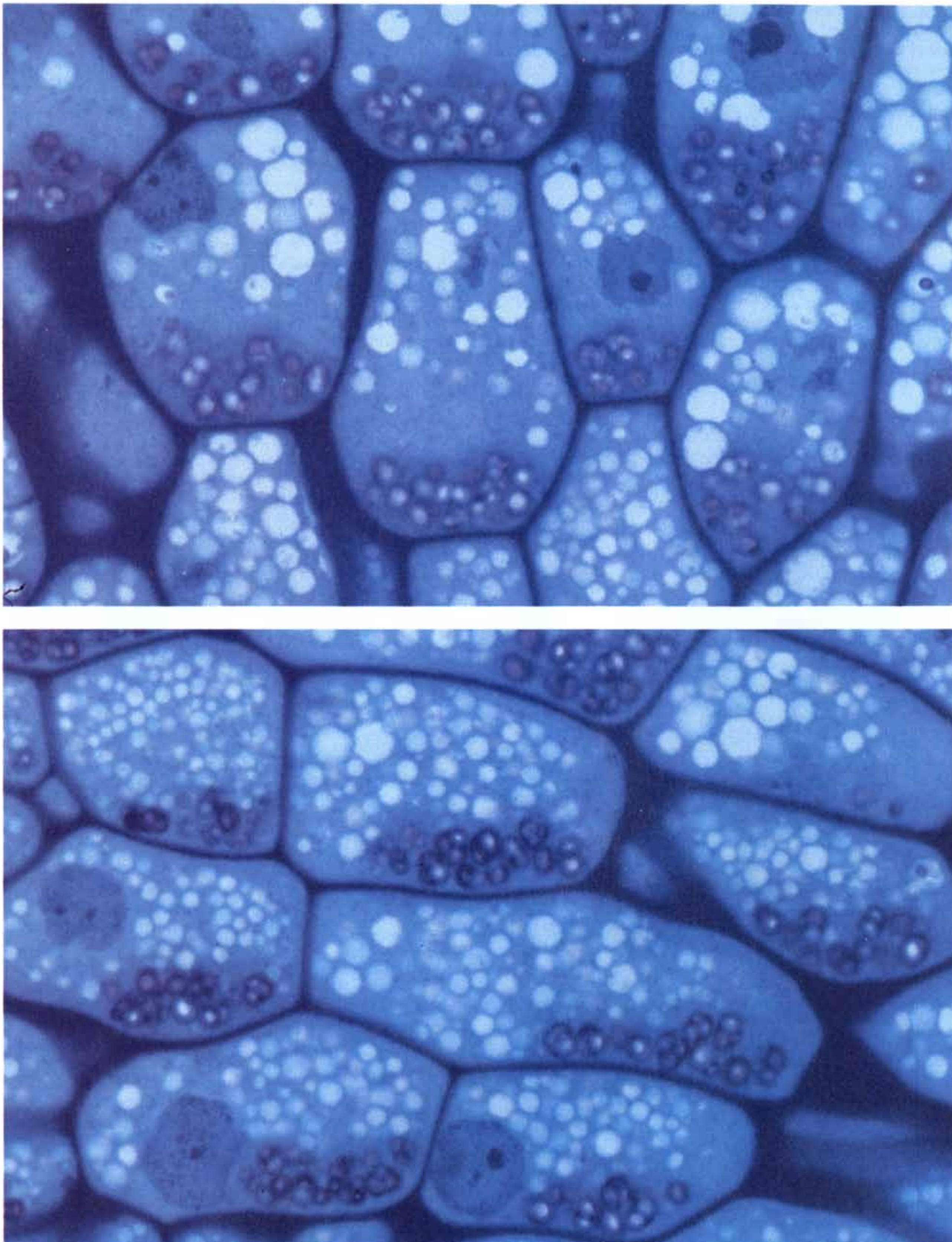
Although the root's precise gravity-perceiving mechanism is not understood, even Darwin recognized that the root cap is the probable site of such perception. The cap, he wrote in 1881, "having the power of directing the movements of the adjoining parts, acts like the brain of one of the lower animals; the brain being seated within the anterior end of the body, receiving impressions from the sense-organs, and directing the several movements." He was among the first investigators to find that removal of the root cap eliminates the ability of the root to respond to gravity. Other investigators have since confirmed the observation and have also shown that replacing an excised cap with one from another root restores gravitropism.

Most workers consider the gravity-detecting portion of the cap to be the central region called the columella, which consists of cells rich in dense amyloplasts: organelles that are filled with starch grains. In vertically oriented roots the amyloplasts reside at the lower end of each columella cell, toward the root tip. Within seconds after roots are gravistimulated, amyloplasts in the columella fall from their former position and rapidly settle along the new lower wall of each cell [*see illustration on opposite page*].

There is substantial evidence that amyloplast displacement constitutes the initial detection of gravity. For example, experimentally treated roots lacking amyloplasts do not respond to gravity. In some cases tilting a plant on its side for a mere 12 seconds tells the plant that it has been reoriented; amyloplasts appear to be the only organelles in the root cap that are significantly displaced within such a short period. In roots made nonresponsive to gravity by the removal of the root cap, recovery of responsiveness to gravity correlates with the formation and settling of new amyloplasts; decapped roots regain gravitropic sensitivity in from 14 to 22 hours, the time it takes for new amyloplasts to form and settle in certain cells near the tip of the decapped root.

If amyloplast settling is in fact the root's gravity-perceiving mechanism, it is probably the only step of the gravitropic response in which gravity directly pulls down some target (in this case the amyloplasts); the other phases seem to be mediated by chemical activity. Before discussing these other phases we should note that a laboratory-developed mutant of one type of plant has been described that lacks amyloplasts in its cap yet still has a nearly normal gravitropic response. This finding suggests that gravitropism can take place in the absence of amyloplasts. It also suggests that amyloplasts may not be the initial gravity detector in roots or that, in the absence of amyloplasts, roots can activate an as yet unknown alternative gravity-detecting mechanism.

Whatever mechanism accounts for the perception of gravity, the ultimate effect of gravity detection—root curvature—clearly results from asymmetric growth in the elongating zone. When a root is oriented vertically, it grows uniformly on all sides. In con-



CELLS in the columella region of the root cap are thought to detect gravity. When the root is vertical (*top*), dense organelles known as amyloplasts (*dark bodies with white grains*) reside along the lower side of the cells. When the root is turned to be horizontal (*bottom*), gravity causes the amyloplasts to settle quickly toward what then becomes the lower side of the cells. Such sedimentation appears to set in motion a series of events that ultimately cause the root to curve downward in the direction of gravity.

trast, when a root is turned horizontally, the upper side grows faster than the lower side; slower growth at the bottom makes the root curve downward.

To discover the factor that directly alters the growth pattern of a gravistimulated root, we and other investigators sought to determine the precise changes in growth along the upper and lower sides of the root. We considered many possible permutations. For instance, does the growth rate along the upper side increase while the rate along the lower side remains constant? Does the rate on the lower side decrease while the rate on the top increases? Does the growth rate on both sides increase, with the top rate increasing more than the bottom, or does the growth rate on both sides decrease, with the bottom rate decreasing more than the top?

Amy J. Nelson, a graduate student working with one of us (Evans), determined the typical growth pattern of gravistimulated corn roots. She put young seedlings in a chamber that had uniformly high humidity, thereby providing the moisture needed for normal root growth. In this experiment, as in the others we shall discuss, each seedling was placed in an environment that eliminated the possible effects of an uneven distribution of moisture, light, nutrients or temperature on the direction of root growth; this ensured that any curvature in the root could be attributed to gravity alone. Nelson positioned the white root of each seedling in front of a white background and put small black beads along the root surface, which is naturally sticky. A television camera focused on the marked root then sent signals to a computer. The computer tracked the gradual movement of the beads as the root grew, calculating the growth rate of various parts of the root before and during the response to gravity.

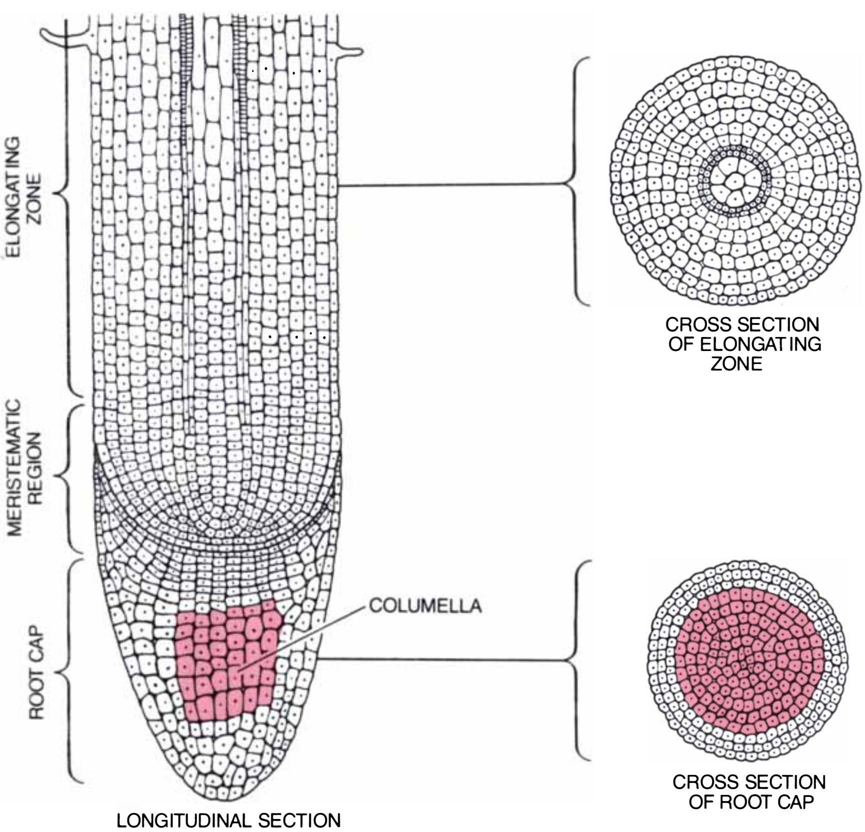
Nelson found that gravistimulation causes a reduction in the growth rate on both the upper and the lower surfaces of the root, but the greatest reduction occurs along the lower surface, from two to three millimeters behind the cap. Other workers employing different test methods with different kinds of plants have similarly found that retarded growth at the lower side of the root causes downward curvature in gravistimulated roots.

What could account for the slowing of growth along the lower side? The growth of plant cells is controlled largely by hormones, which suggests that there is an increase in the concentration of some growth-inhibiting hormone along the lower side of the elongating zone. One hypothesis for root gravitropism suggests that the growth inhibitor in question is the hormone abscisic acid, which forms naturally in plant roots. When abscisic acid is applied in large doses to the outside of roots, it is absorbed and inhibits growth. In roots of certain corn seedlings that do not exhibit gravitropism in the dark, illuminating the root cap induces gravitropism within an hour, during which time the rate of synthesis of abscisic acid in the root cap has also been found to increase.

In spite of these findings, more recent evidence has led us to conclude that the abscisic acid hypothesis is incorrect. One of us (Evans) has shown that the hormone inhibits root elongation only when it is applied to roots at concentrations significantly higher than are thought to occur naturally.

What is perhaps more significant, another of us (Moore) has found that roots of corn seedlings grown in the presence of an inhibitor of abscisic acid synthesis have undetectable levels of abscisic acid but nonetheless curve downward in response to gravistimulation. Similarly, roots of a corn mutant not capable of synthesizing abscisic acid respond to gravity. Moreover, one of us (Evans) has found that roots immersed in a high concentration of abscisic acid curve downward after being gravistimulated. The latter finding is significant because it is unlikely that a subtle gravity-induced gradient of the hormone could by itself induce curvature when the entire root is immersed in saturating levels of the substance.

Strong evidence suggests that the hormone auxin controls gravitropic root curvature. Auxin, which like abscisic acid occurs naturally in roots, is a powerful inhibitor of their growth, even at concentrations from 100 to 1,000 times lower than those at which abscisic acid is effective. In roots made nonresponsive to gravity (by removal of the root cap), auxin applied in a small dose to one side of the elongating zone can induce the roots to curve toward the side where the auxin has been applied—a finding that suggests a subtle, physiological increase of auxin at the lower side of a horizontal root should be sufficient to cause downward curvature.



TERMINAL PARTS of the root are where gravitropism takes place. The root cap contains the gravity-detecting columella cells in its core. The elongating (or growing) zone, which frequently reaches a length of about five to six millimeters, is the site of gravitropic curvature. The intervening meristematic region is where new root cells are produced.

After radioactively labeled auxin was applied uniformly to the elongating zone of a gravistimulated root, the labeled hormone moved toward the lower side of the root, suggesting that at least some of the naturally occurring auxin moves in such a way that it collects at the bottom of the elongating zone. Moreover, chemicals that interfere with the cell-to-cell movement of auxin in roots, thereby presumably impeding the increased deposition of auxin in the usual area of gravitropic curvature, overcome the root's ability to curve in response to gravity.

To demonstrate more conclusively that auxin accounts for downward curvature in gravistimulated roots, it would be necessary to show directly that natural auxin within the root becomes concentrated at the bottom of the elongating zone just before the onset of gravitropic curvature. Such evidence has not yet been found.

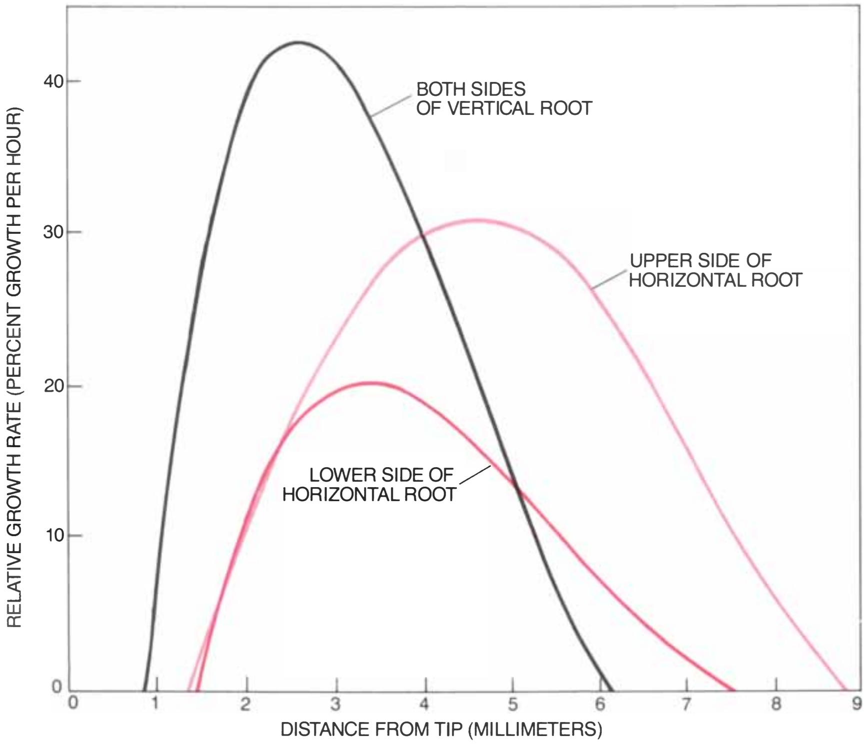
If the settling of amyloplasts in the columella of the root cap accounts for the perception of gravity, and if an increase of auxin at the lower side of the elongating zone accounts for the ultimate response of the root, what ties these events together? That is, what constitutes the mysterious transduction phase of gravitropism?

On the basis of recently collected data we propose that the settling of amyloplasts triggers the release of calcium ions (Ca^{++}) from organelles along the lower side of the columella cells. In turn the released calcium activates transport systems that move calcium and auxin downward from cell to cell toward the lower side of the root cap. The calcium at the bottom of the cap then facilitates the movement of auxin from the cap back along the lower side of the elongating zone.

We first suspected that the release and movement of stored calcium might be an important initial step in the transduction phase of root gravitropism after several papers published in the 1960's and 1970's reported that calcium appears to move from cell to cell when shoots respond to gravity. These experiments showed that calcium migrates toward the top surface of a gravistimulated shoot before the shoot curves upward.

To test our idea Konrad M. Kuzmanoff and Timothy J. Mulkey, graduate students working for one of us (Evans), applied EDTA, a chemical that binds and immobilizes calcium, to the caps of corn roots. To our surprise, the EDTA-treated roots, which continued to grow at their normal rate, became totally unresponsive to gravity. Removing EDTA and replacing it with cal-

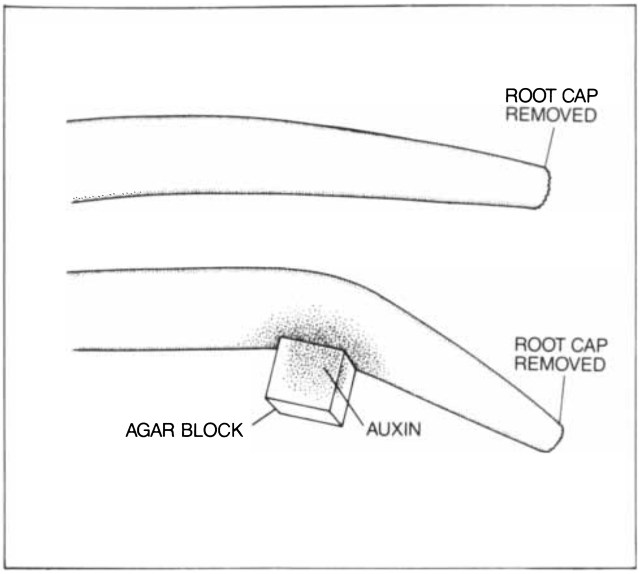
GRAVISTIMULATED CORN ROOTS lacking a cap (*top*) and having an intact cap (*bottom*) demonstrate the importance of the cap to root gravitropism. If the cap is removed, the root grows normally but no longer curves downward when it is oriented horizontally (gravistimulated). In contrast, a root with an intact cap does curve when it is reoriented.



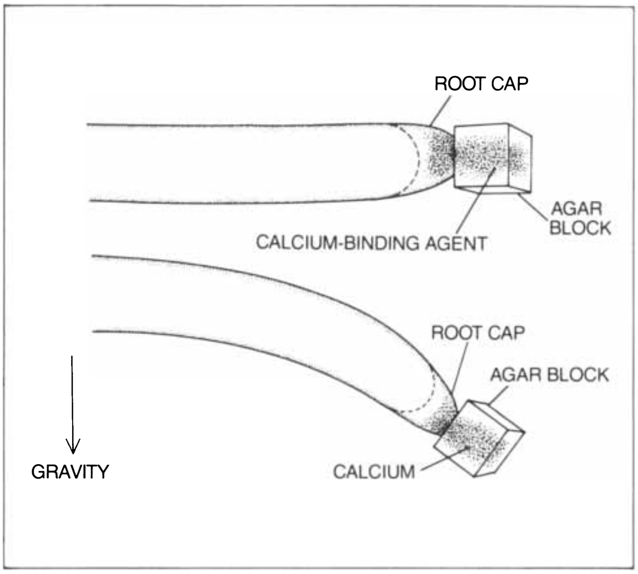
COMPARISON OF GROWTH RATES at different positions along a vertically oriented and a horizontally oriented corn root reveals part of the mechanism by which roots respond to gravity. Vertical roots grow symmetrically in the elongating zone, whereas horizontal roots grow asymmetrically and at a reduced rate. In horizontal roots the lower side grows more slowly than the upper one, so that the root curves downward. One model of gravitropism suggests that this asymmetric reduction in growth results from an uneven redistribution of auxin, a plant hormone that is known to inhibit root growth. This model further suggests that auxin movement may be controlled by calcium ions in the root cap.

cium restored the roots' gravitropic responsiveness. Kuzmanoff and Mulkey hence provided strong evidence that free and mobile calcium in the root cap is essential for root gravitropism. Their observation has since been strengthened by other studies showing that a variety of treatments preventing calcium movement can all abolish gravitropism.

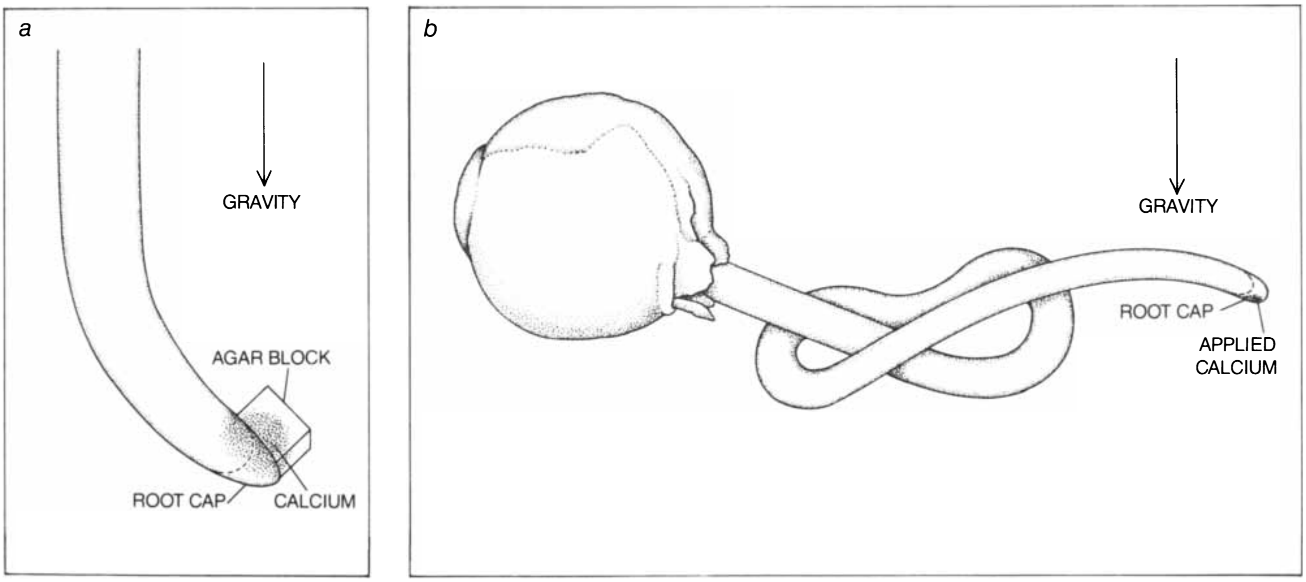
Taking up where Kuzmanoff and Mulkey left off, June S. Lee, who at the time was visiting Ohio State University from Kang Reung National University in Korea, determined the probable direction of calcium movement within gravistimulated roots. First he studied roots that were not gravistimulated and discovered that calcium has a powerful influence on their direction of growth. When he applied calcium to the root cap unevenly, the root curved toward the area of greatest calcium concentration. In fact, when calcium was applied continuously to one side of the cap of a vertically oriented root, the root eventually curved through a complete 360-degree loop. Knowing that gravistimulated roots curve downward, Lee concluded



DECAPPED ROOTS from corn seedlings help to demonstrate that an uneven distribution of auxin could be responsible for asymmetric growth in the elongating zone of gravistimulated roots. Removal of the cap makes roots nonresponsive to gravity (*top*), but when auxin in a hormone-filled agar block is applied to one side of a decapped root (*bottom*), the root grows more slowly along that side and finally curves toward the applied hormone.



CORN ROOT to which a calcium-binding agent was applied at the cap grows horizontally when it is turned sideways (*top*) because the binding agent prevents calcium from being redistributed in response to gravity. Replacing the agar block containing the binding agent with one containing calcium restores the root's gravitropism (*bottom*). Such results indicate that mobile calcium in the cap is essential to the root's ability to respond to gravity.



CORN ROOTS to which calcium was applied curve toward the side of application. Soon after an agar block containing calcium was placed on the right side of the cap of a vertical root (*a*) the root curved to the right. Similarly, long-term experiments in which the location of the calcium applied to the root cap was periodically changed (*b*) generated several curves, each developing as the root arched toward the calcium. These results indicate that an unequal distribution of calcium in the cap of a gravistimulated root—and particularly an excess along the lower side of the root cap—contributes to downward curvature in the elongating zone.

that such curvature may result from an excess of calcium at the bottom of the cap, perhaps deposited by the downward movement of calcium.

Next Lee showed that calcium could in fact move downward within the cap. He applied radioactive calcium evenly across corn roots, oriented the roots either vertically or horizontally and measured the movement of the radioactive ions. In vertically oriented roots the radioactive calcium remained uniformly distributed. In horizontally oriented roots, on the other hand, the radioactive calcium moved toward the lower side of the root. This directional movement of calcium was particularly strong in the root cap.

Results from a completely different kind of study lend support to the proposal that calcium moves downward in gravistimulated roots. By placing special microelectrodes close to a vertically oriented root of cress, H. M. Behrens and his co-workers at the University of Bonn showed that an electric current flowed in a symmetrical pattern along the root surface and into the root near the tip. When the workers put the root in a horizontal position, they found that the current pattern became asymmetric; in particular, the current along the upper side of the cap flowed out through the top, whereas the current along the lower side flowed into the cap. They found evidence that the current was carried by a flow of hydrogen ions (H^+).

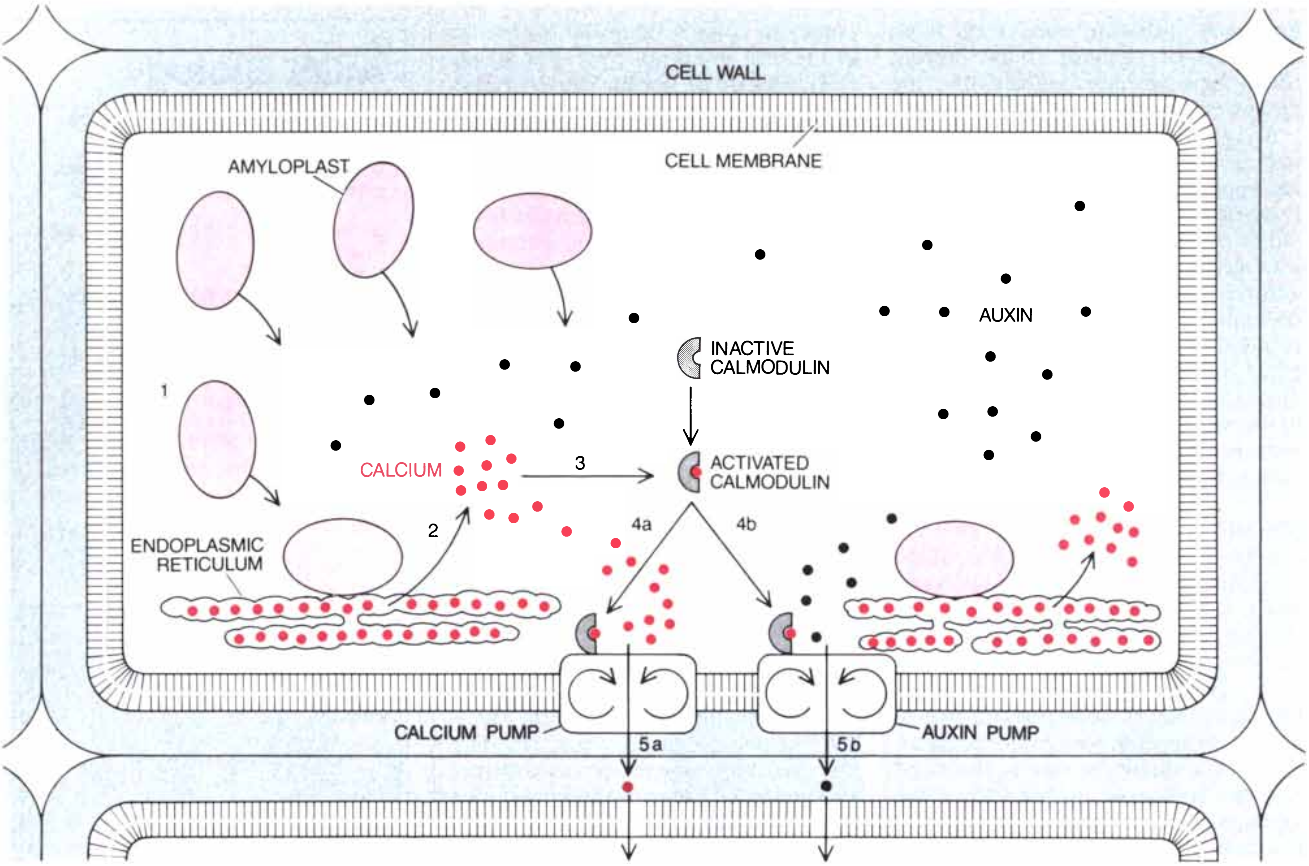
Thomas Björkman and A. Carl Leopold of Cornell University recently confirmed these observations in studies of corn roots. They found that from two to six minutes after the roots are oriented horizontally the direction of the current in the root cap changes: it moves upward and out through the top of the cap. Björkman and Leopold speculate that the upward flow of hydrogen ions could indirectly reflect a flow of calcium ions to the lower side of the cap; to maintain electrical neutrality, the cap would have to balance such calcium movement with a counterflow of other positive charges, such as hydrogen ions. (The amount of upward current measured by Björkman and Leopold is consistent with our estimates of the calcium that moves to the bottom of the root cap.) If the upward current does in fact reflect calcium migration, the timing of calcium movement suggests that the migration results from amyloplast settling; the calcium movement immediately follows settling, and it precedes downward curvature of the root.

Just how amyloplast settling triggers calcium movement and how calcium comes to collect along the lower side of the root cap is not clear, but Barbara G. Pickard of Washington University in St. Louis has proposed a reasonable hypothesis, one that is consistent with the results of Björkman and Leopold. In columella cells, amyloplasts displaced by gravistimulation fall onto the endoplasmic reticulum, a complex of calcium-rich membranes and vesicles. Pickard suggests that the pressure of the amyloplasts on the endoplasmic reticulum causes calcium ions to escape from the complex, resulting in a localized elevation of the calcium level along the lower side of the cells.

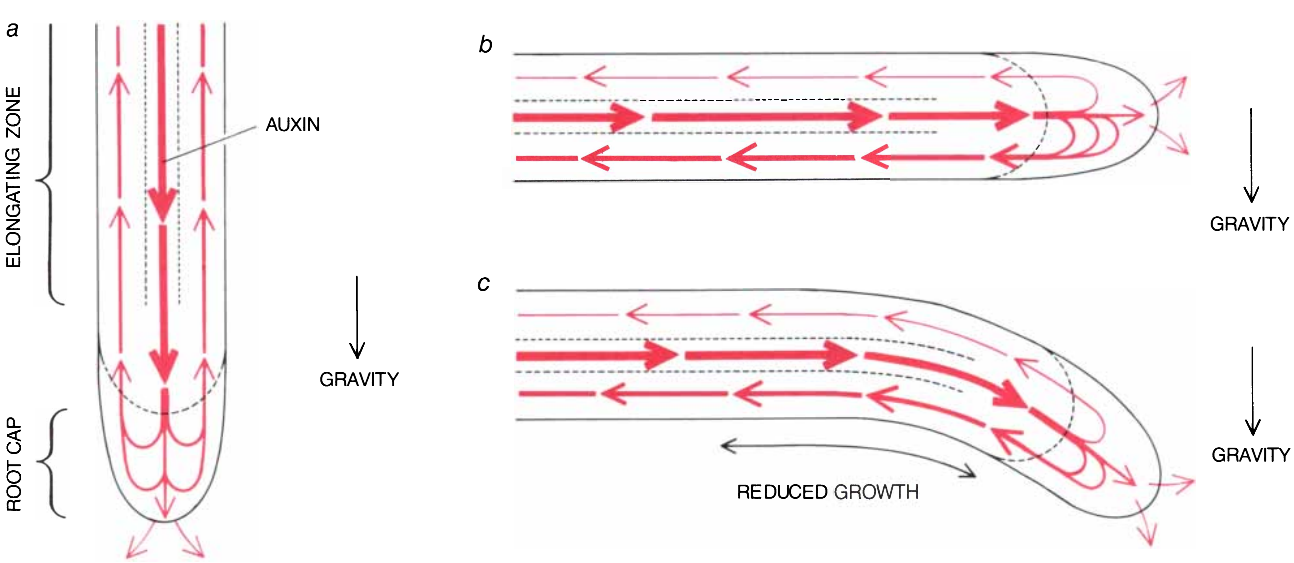
We propose an addition to this model. We suspect that when the calcium level reaches a certain threshold, the ions activate calmodulin: a small protein known to be a powerful activator of many enzymes important to cellular function, not only in plants but also in animals and even in some microorganisms. Once it is activated by calcium, calmodulin appears to stimulate calcium movement by "turning on" calcium pumps, which are large enzymes thought to traverse the cell membrane. The activated pumps at the lower side of the cells then excrete excess calcium, which eventually accumulates along the lower side of the root cap.

The calmodulin required by this model does appear to be available in the root cap. Charles L. Stinemetz, a graduate student working with one of us (Evans), verified its presence in corn roots and showed that the concentration of calmodulin is four times as great in the cap as it is in the elongating zone. Lending further credence to the potential role of calmodulin, Stinemetz also found that root-cap calmodulin may be necessary for root gravitropism. When he applied calmodulin inhibitors to the caps of roots, the inhibitors strongly retarded gravitropic curvature.

In order to explain the remaining steps of the gravitropic response—the downward movement of auxin in the root cap and the calcium-enhanced movement of auxin to the lower side of the elongating zone—we must first describe what is thought to be the usual pattern of auxin movement through the root. In roots that are vertically oriented auxin moves toward the root cap through the core of the elongating zone. Some of the auxin traveling into the cap is either metabolized or moves through the tip into the growth medium. At the same time some auxin branches off toward the sides of the cap, where it is redirected up into the elongating zone by cells along the outer margins of the root [*see bottom illustration on next page*]. The symmetry of



PROPOSED MECHANISM linking amyloplast settling to the downward movement of calcium ions (*red dots*) and auxin (*black dots*) in the cap of gravistimulated roots depends on the enzyme calmodulin, an important activator of enzymes in plant and animal cells. When a root is turned on its side, amyloplasts in columella cells fall toward the lower side of each cell, making contact with the endoplasmic reticulum, a complex of calcium-rich membranes (*1*). The resulting pressure on the endoplasmic reticulum induces the membrane complex to release calcium into the surrounding cytoplasm (*2*). When the calcium level in the cytoplasm along the lower side of the cell reaches a certain threshold, the calcium binds to and activates calmodulin in the area (*3*), which in turn activates two types of enzyme in the cell membrane along the lower side of the cell: a calcium pump (*4a*) and an auxin pump (*4b*). Calcium pumps transport excess calcium into the cell wall (*5a*) and auxin pumps move excess auxin into the wall (*5b*). By different mechanisms, the calcium and the auxin then migrate through the underlying cells toward the lower side of the root cap.



PATTERN OF AUXIN MOVEMENT (*arrows*) in a root changes when the root is gravistimulated. In a vertical root (*a*) auxin travels from the elongating zone toward the root cap through the central part of the root. After the auxin enters the cap some of the hormone is thought to move across the length of the cap and out through the tip; at the same time some of it branches off to the sides and then symmetrically flows back into the elongating zone. After the root is turned horizontally (*b*) the pattern of auxin flow becomes asymmetric. The authors propose that much of the auxin entering the cap moves toward the lower side of the cap, where collected calcium somehow increases the rate at which the auxin travels back into the elongating zone. The resulting auxin excess within cells along the lower half of the elongating zone then inhibits its growth there, leading to downward curvature of the root (*c*).

this movement causes the vertical root to grow straight.

In gravistimulated roots the pattern of auxin movement in the root cap changes. On the basis of recent data collected by one of us (Hasenstein), we propose that calcium-activated calmodulin in the columella cells activates not only calcium pumps in the lowermost part of the cell membrane but also auxin pumps. Such auxin pumps are known to exist in plant-cell membranes. We think these pumps transport auxin out through the lower side of the columella cells, so that much of the auxin entering the root cap from the elongating zone travels downward to the lower side of the cap instead of dividing symmetrically between the upper and lower sides. We further propose that the elevated concentration of calcium in cells along the lower side of the root cap then somehow enhances the rate at which auxin in the lower part of the root cap travels back into the lower side of the elongating zone.

We shall now return to the question posed more than a century ago: By what mechanism does a horizontally oriented root respond to gravity? We suggest that gravity pulls on amyloplasts in the columella cells of the root cap. The amyloplasts, falling onto the endoplasmic reticulum at the lower side of the cells, cause calcium levels in the cytoplasm to rise, thereby activating calmodulin, which in turn activates calcium and auxin pumps in the membrane at the lower side of the cells. The pumps force calcium and auxin out through the lower side of the cells, leading ultimately to an accumulation of calcium and auxin at the lower side of the root cap. The elevated concentration of calcium in this part of the cap then enhances the loading of auxin into the pathway that carries auxin back toward the lower side of the elongating zone. There the auxin markedly inhibits growth and causes the root to curve downward in the direction of gravity.

It is not surprising that the downward pattern of growth produced by gravitropism almost always leads the root toward a supply of moisture and nutrients. Our recent findings and those of other botanists confirm Darwin's suggestion that this self-preserving behavior of the root (or, in his words, the "radicle") is controlled by the cap. Indeed, investigators of root gravitropism, and of other root behavior, have reason today to echo Darwin's 1881 assessment that "there is no structure in plants more wonderful, as far as its functions are concerned, than the tip of the radicle."

Human-powered Watercraft

In striving for ever higher speeds the familiar racing shells propelled by eight oarsmen may have to give way to unconventional watercraft. Such a record-setting vehicle was designed and built by the authors

by Alec N. Brooks, Allan V. Abbott and David Gordon Wilson

Until recently the slender shells, or racing rowboats, made familiar by the Olympic Games, the races on the Thames River in England and other rowing regattas held throughout the world, were the fastest human-powered watercraft. The fastest of these, powered by a crew of eight oarsmen, achieve speeds of 12 knots over a standard 2,000-meter course. (One knot is equal to about half a meter per second.) Human-powered watercraft that are not bound by the arbitrary restrictions of officially sanctioned rowing events are likely to equal or surpass this level of performance. Designers of these unconventional craft are dispensing with oars and taking full advantage of modern high-efficiency propellers. They are even dispensing with hulls as they explore innovative ways to reduce the resistance against motion, called drag, that water exerts on a moving boat.

Indeed, two of us (Brooks and Abbott) have developed just such a record-setting human-powered watercraft. The craft, *Flying Fish II,* is ridden like a bicycle. It has a pair of hydrofoils, or underwater wings, and a high-efficiency propeller. It enables a single rider to complete a 2,000-meter course significantly faster than a single rower in a shell can, and it has attained a maximum speed of 13 knots over short distances.

Regardless of its design—whether it is a crude flotation device propelled by underwater kicking, a wood raft pushed along by poles, a dugout canoe powered by paddles or a dinghy moved forward by sweeping oars—every watercraft must contend with four basic forces: weight, lift, thrust and drag. Weight and lift are the simplest forces to understand. Weight is simply the gravitational force pulling down on the craft and its occupants; lift is the force that acts upward, counteracting the weight. As long as a boat does not experience any vertical acceleration, lift is equal to weight.

For most watercraft lift is generated by buoyancy: the displacement of water by the craft's hull. The lift is equal to the weight of the water displaced, and it operates even in the absence of motion. In addition many high-speed boats take advantage of dynamic lift, which is produced as the boat moves through water. A common example of dynamic lift is planing: when the bottom of the hull continuously deflects water downward so that lift is produced as a reaction force. A boat that relies on planing for most of its lift rides higher in the water—often right at the surface—and requires less buoyancy. Until recently designers of human-powered watercraft had not been able to successfully incorporate dynamic lift into their vehicles.

Thrust is the force (produced by the actions of the operator in the case of human-powered watercraft) that propels the craft. Drag is the force that by definition acts in the direction opposite to the direction of the craft's motion. If a boat is moving at a steady speed, the thrust is equal to the drag. In summary, at constant speed lift balances weight and thrust balances drag.

In order to translate efficiently a given human power input into speed the most important objective is to minimize drag. One obvious way to lessen drag is to reduce the weight of the boat. Once a boat begins to move, its source of lift almost always exacts a drag penalty. By minimizing the weight of the boat the required lift is reduced, and hence the drag associated with the lift is lessened. Since the craft's operators are not likely to be overweight (assuming they are healthy, athletic individuals to begin with), the weight reduction must apply primarily to the vehicle itself.

Efforts to this end have led to racing shells that weigh only a small fraction of the operator's weight—a relation similar to that of a modern racing bicycle and its rider. In the past shells were generally made of cedar, spruce and mahogany, and they were made lighter by thinning their hulls. (Indeed, the term "shell" arose because a careless finger could easily puncture a wood hull.) In the 1950's experimental shells that had a skin of glass-fiber-reinforced plastic were tried, and by the end of the 1960's commercially available composite-based boats had challenged the dominant position of wood boats in rowing circles. Today the wood shell is becoming a rarity. Sophisticated composite materials consisting of a resin matrix interlaced with fibers of a polymer or graphite have brought down the weight of the lightest single-person shell to less than 10 kilograms.

Assuming that the weight of the racing shell has been reduced to its practical minimum, a designer's attention must turn to other ways of minimizing drag. Shells have what are called displacement hulls: virtually all their lift is produced by the buoyancy of the hull. Displacement hulls have the unique property that their drag approaches zero as their speed through the water approaches zero. Hence at very low speeds displacement-hull vehicles have extremely low drag and

AMONG THE FASTEST human-powered watercraft are conventional racing shells and the authors' unconventional vehicle, *Flying Fish II*. The shell, shown here being rowed by four-time Olympian John Van Blom, was built by Alfred Stämpfli AG of Switzerland. *Flying Fish II,* ridden by one of the authors (Abbott), is powered by a pedal-driven propeller and supported by two hydrofoils, or underwater wings, while racing. The floats are actually above the water surface and are meant to support the craft only at low speeds.

are among the most efficient of all vehicles. Racing shells, however, do not operate at low speeds.

As a shell's speed increases, its drag increases dramatically owing in part to the formation of waves that emanate from the bow and stern. The energy needed to produce these waves is manifested as wave drag. Wave drag increases rapidly with increasing speed but in an uneven fashion because the bow wave can interact constructively with the stern wave (so that the waves are in phase and reinforce one another) or destructively (so that the waves are out of phase and tend to cancel one another) as the craft picks up speed. At a speed called hull speed the bow is at the crest and the stern is at the trough of a single wave; in its passage through the water the hull has literally created a hill of water through which the boat must push. At this point a great expenditure of power is needed to increase the boat's speed. The human power plant cannot supply the required effort, and so hull speed acts as the effective speed limit of a human-powered displacement-hull vehicle.

Hull speed is proportional to the square root of the waterline length of a boat. Human-powered watercraft that have long displacement hulls are therefore less hindered by wave drag than boats that have short hulls with the same overall buoyancy. On the other hand, for a given buoyancy long, slender hulls have more wetted surface area than short, wide hulls. The greater the wetted surface area, the greater the drag caused by the friction of the water as it flows past the surface of the hull. This type of drag is known as skin-friction drag. Hence as a boat is made more slender, wave drag diminishes but skin-friction drag then becomes more of a problem.

A hull designed for speedy boats must therefore be shaped to minimize the sum of wave and skin-friction drag. Shells are designed to compete in six-to-seven-minute races at power levels of about half a horsepower per rower. (One horsepower is equal to approximately 750 watts.) The resulting optimal length-to-width ratios of these sleek craft exceed 30. A single-person shell, for example, has a length of between eight and nine meters and a width of no more than 30 centimeters. It turns out that the optimal shell shape results in a skewed distribution of drag at racing speeds: 80 percent of the drag operating on the shell is due to skin friction and 20 percent is due to wave production.

Given that skin friction is the dominant source of drag operating on a shell at racing speed, a substantial reduction in drag is possible if skin friction can be reduced. Skin friction arises from a thin layer of water, known as the boundary layer, that flows past the boat's hull. There are two fundamental types of boundary layer: laminar, in which the flow is smooth and steady, and turbulent, in which the flow is chaotic and unsteady. Laminar boundary layers produce much less skin-friction drag than turbulent boundary layers do. The boundary layer on a shell is laminar at the bow, but only a short distance back from the bow it typically undergoes a transition to turbulent flow. Drag is significantly reduced if the transition is delayed, thereby increasing the area of laminar flow on the hull.

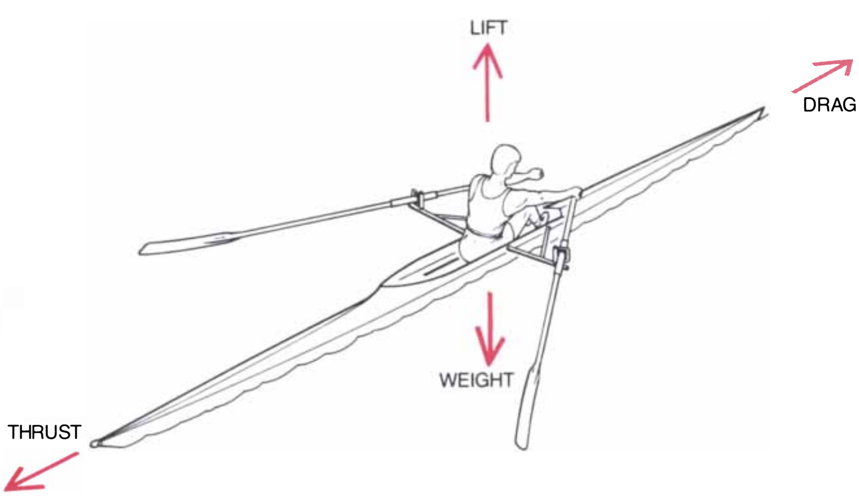
One method of extending the laminar boundary layer that is applied in some specialized underwater vehicles is the injection of long-chain polymers (sometimes referred to as slippery water) into the boundary layer near the front of the craft. Race-sanctioning organizations are not likely to allow this practice in competition, if for no other reason than that it pollutes the water. A similar approach that might be allowed, however, would entail carefully cultivating a layer of naturally slimy algae or some other innocuous microorganisms on the hull.

Boundary-layer suction is another technique that has been applied to stabilize a laminar boundary layer. In this approach fluid in the boundary layer is continuously "sucked" into the boat through pores or small slots in the hull surface. Shells could make use of boundary-layer suction if they were outfitted with a porous hull that would allow water to seep in slowly. A small pump would serve to bail the water out occasionally.

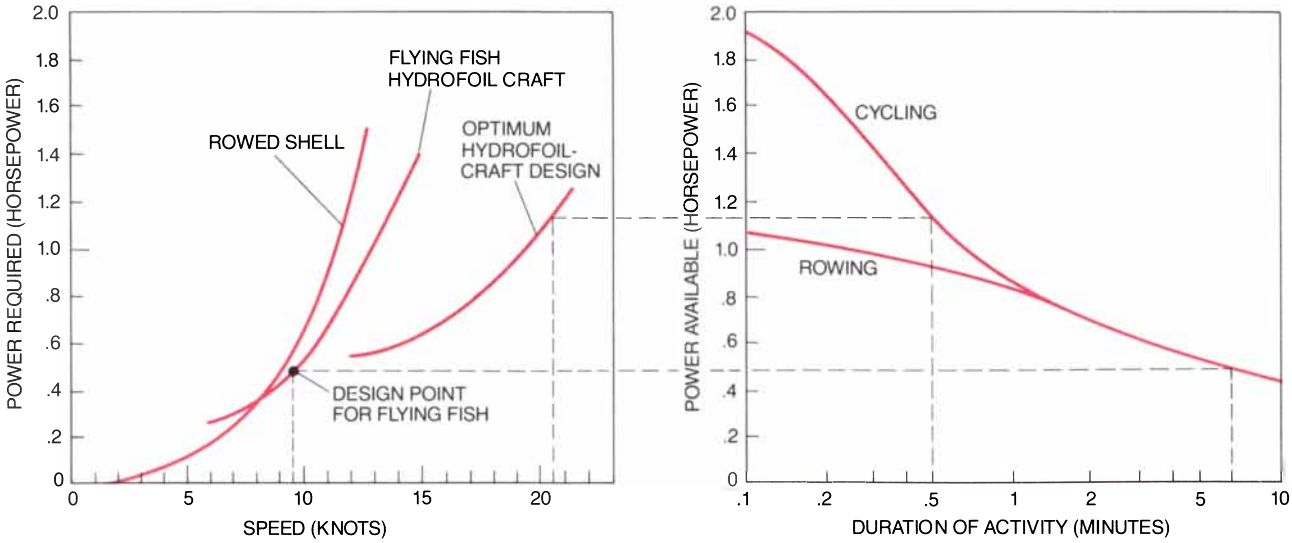
The texture of the wetted hull surface can also play a role in the reduction of skin-friction drag. Investigations under the auspices of the National Aeronautics and Space Administration have shown that a slick waxed surface does not always result in minimal skin-friction drag. Surfaces with very fine grooves running in the flow direction, called riblets, have shown 6 percent less drag than smooth surfaces.

Riblets have been tested on rowed shells by a group from the Flight Research Institute headed by Douglas McLean of the Boeing Company. The group covered a single-person shell with an experimental plastic skin in which grooves had been formed. The spacing between the grooves was three-thousandths of an inch (about 80 micrometers)—finer than the groove spacing on phonograph records. Tests indicated that the shell's maximum speed was increased by 2 percent. Although this may seem like an insignificant amount, it is equivalent to a four-boat-length advantage over a standard 2,000-meter race.

On the basis of such encouraging results the experimental skin was applied to the hull of the U.S. Olympic



FOUR BASIC FORCES must be considered in designing boats such as the racing shell shown here: weight, lift, thrust and drag. Weight is the gravitational force acting on the boat and the operator. Lift is normally generated by buoyancy, the upward force equal to the weight of the water displaced by the boat's hull. Additional lift, called dynamic lift, can be produced by the flow of water under the hull. Thrust, in the case of human-powered water vehicles, is the force produced by the actions of the operator (here seen rowing) that propels the craft forward. Drag is the resistance to the boat's forward motion; it arises in most craft from the creation of a wake (wave drag) and the friction between the hull and the water flowing past it (skin-friction drag). When a boat has a constant speed, lift balances weight and thrust balances drag. The key objective in boat design is to minimize drag at the normal operating speed of the boat. At the speeds necessary for competitive rowing, drag is minimized by making a shell light, long and narrow.



POWER LEVEL necessary for a human-powered watercraft to reach a certain speed for a certain length of time depends on the craft's design. The graph at the left shows the power required for a rowing shell, which relies on the displacement of water by its hull for most of its lift, compared with two other craft designs (one being the authors'), which rely on the dynamic lift produced by hydrofoils. At low speeds displacement-hull craft are more efficient than hydrofoil craft. Actually hydrofoil craft have a minimum speed below which the hydrofoils cannot support the combined weight of the craft and the operator. At higher speeds, however, hydrofoil vessels are more efficient than displacement-hull craft. The graph at the right shows how the power a champion athlete can supply diminishes with the effort's duration. For brief durations the power output generated by the cycling motion is considerably higher than the output generated by the rowing motion. An optimal hydrofoil design could make it possible to reach speeds of more than 20 knots. Such a feat would require power levels that can be achieved only by cycling and only for a few seconds. A craft incorporating such a hydrofoil would be difficult to get started: its "takeoff" speed would be more than 11 knots.

team men's coxed-four rowing shell. (A coxed boat is steered by a coxswain, who does not row but calls out the rowing cadence.) The team made an excellent showing, winning the silver medal in the 1984 summer games.

In addition to low drag, another essential ingredient for a successful racing shell is good propulsive efficiency: as much as possible of the power from humans must be converted into useful thrust. In the case of rowing, two major advances in propulsive efficiency date from the mid-19th century. One was the development of the modern rigger in 1843. The rigger is a tripodlike device attached to the side of the boat. The oarlock, or pivot point for the oar, is located at the apex of the tripod. Since the oarlocks no longer needed to be attached directly to the gunwales, or edges of the sides of the boat, the hull could be narrower (reducing wave drag) and the oars could be longer (enabling rowers to take longer and more efficient strokes).

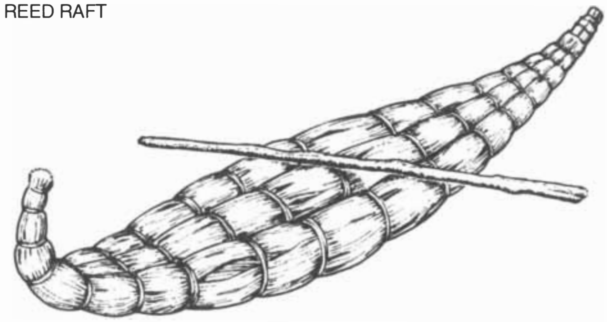
The second advance was made in 1856· the sliding seat. Until that time rowed boats were propelled through the use of the muscles of the arm, shoulders and back, whereas the larger muscles of the legs were used only to brace or support the body. The motion when rowing was one of heavy straining against a slowly moving resistance. The sliding-seat arrangement allows the energy of the leg muscles to be harnessed as the seat moves fore and aft with the bending and straightening of the legs during the rowing cycle. The first sliding seat was a rather crude device consisting of a sheepskin pad sliding on a greased panel. The sliding seat on bearings, still in use today, was invented in the U.S. in 1857.

A shell with a variation of the sliding seat was rowed by Peter Michael Kolbe to win the 1981 world championships in Munich. In contrast to conventional shells, which have sliding seats and fixed riggers and stretchers, or footboards, Kolbe's custom shell was equipped with a fixed seat and a sliding frame that supported the riggers and stretchers. Under this arrangement the rowing motion is the same as the motion for conventional shells, but since most of the rower's mass is on the fixed (rather than sliding) seat, the oscillations of the center of mass (which more or less coincides with the rower) are greatly diminished. This in turn diminishes the oscillations in velocity that a shell is subject to as it travels through the water. (These oscillations are manifested in a conventional shell by its distinctive jerky motion when it is rowed forcefully.)

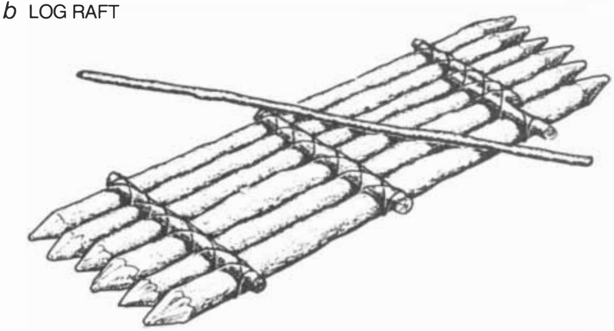
Because skin-friction drag is not a linear function of the velocity of the water in relation to the hull (it is in fact proportional to the square of the velocity), a fluctuating speed always produces more drag than would occur if the boat moved steadily at the average speed. The drag reduction obtained by the sliding-rigger arrangement is only slight, but it is enough to make a significant difference in the racing world. In the 1982 world championships five boats in the men's finals had fixed seats and sliding riggers. In 1983 all six finalists used sliding-rigger boats. After 1983, however, sliding-rigger boats were ruled ineligible for competition.

Although the addition of the rigger and the sliding seat significantly augmented the propulsive efficiency of rowing, rowing nonetheless has a fundamental limitation. Oars and paddles are basically drag devices: they generate thrust by slipping backward through the water. The slippage represents an efficiency loss; it can be reduced by increasing the size of the oar blade, but only to a limited degree because of practical constraints. Moreover, the aerodynamic drag caused by the blades when they are out of the water during the return stroke can be quite significant, particularly under windy conditions.

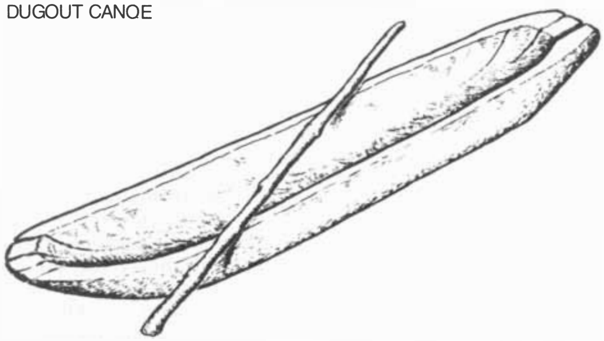
The efficiency of a propulsion system is defined as the ratio of useful power output, which is the product of the average thrust and the velocity, to the human power input. The detailed physics of rowing is not entirely understood but analysis by many investigators has put the propulsive efficiency of rowing at between 65 and 75 percent. Hence about two-thirds of the *a* REED RAFT



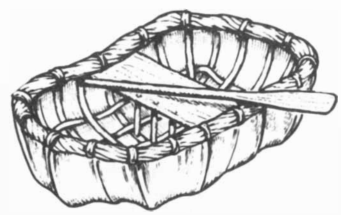
b LOG RAFT



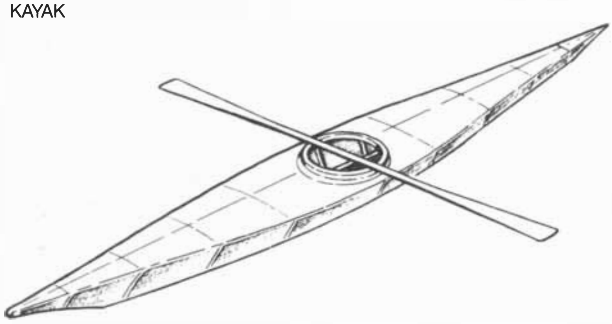
c DUGOUT CANOE



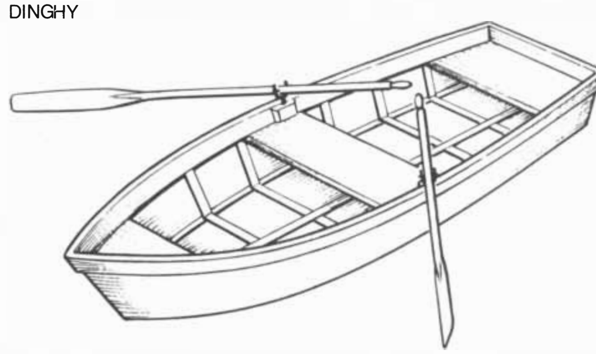
d CORACLE



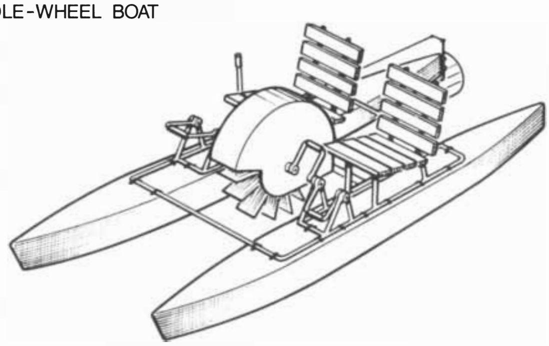
e KAYAK



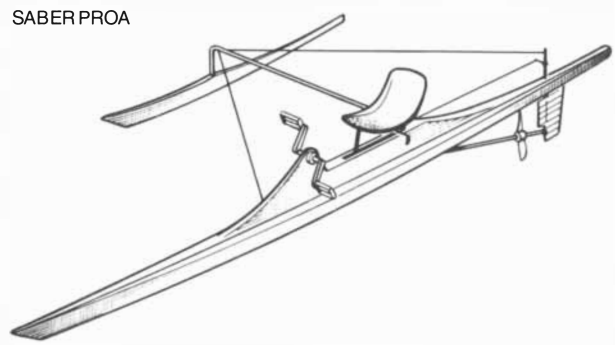
f DINGHY



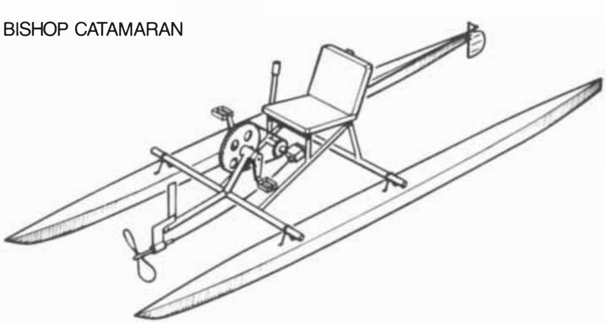
g PADDLE-WHEEL BOAT



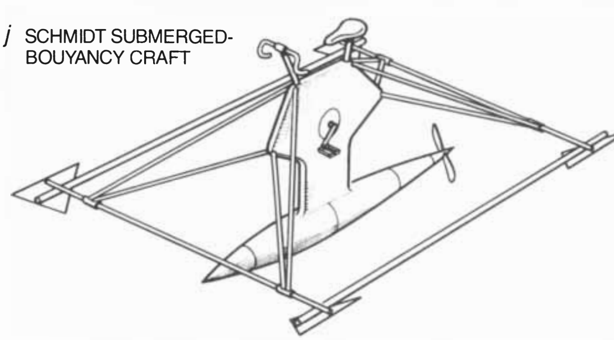
h SABER PROA



i BISHOP CATAMARAN



j SCHMIDT SUBMERGED-BOUYANCY CRAFT



rower's power output is delivered to the boat as useful work; the rest goes into creating disturbances in the water and air.

Motorized boats abandoned propulsive drag devices, such as paddle wheels, long ago in favor of propellers. Interestingly, before the development of small gasoline-fueled outboard motors at the turn of the century, human-powered propeller-driven boats were in fact being developed for practical transportation purposes. Such propeller-driven boats proved to be much faster and less tiring than canoes or rowboats. In the 1890's a three-rider propeller-driven catamaran (a twin-hulled boat) was shown to be 13 percent faster than a three-oarsmen shell over a 163-kilometer course on the Thames River.

Some disadvantages of propellers are that they can be fouled with weeds and can strike bottom in shallow water, but otherwise propellers are particularly suited for human-power applications [see "The Screw Propeller," by E. Eugene Larrabee; SCIENTIFIC AMERICAN, July, 1980]. Slender-bladed, high-efficiency propeller designs can be applied, since the power level is quite low. In addition, propeller-tip speeds are low enough so that cavitation is not a problem. (Cavitation, the formation of bubbles of water vapor, arises when the absolute pressure on some part of the rotating propeller is reduced below the water's vapor pressure; it reduces efficiency and can cause excessive wear on the blade surfaces.) Several new human-powered watercraft have been outfitted with propellers whose efficiencies exceed 90 percent.

The rotating motion of propellers also makes it relatively easy to drive them by an arrangement of pedals, sprockets and chains much like that of bicycles. Such an arrangement takes advantage of the rapid and strong movements of the legs. The circular pedaling action in bicycling remains the most efficient practical method of transferring continuous power from a human being to a machine. (It is not coincidence that record-setting human-powered air and land vehicles depend on a bicyclelike drive train.)

A champion cyclist can produce nearly two horsepower for a few seconds of maximum effort. For periods of continuous exertion lasting for six minutes, however, the power output is generally no more than half a horsepower. Various factors affect power production, including pedaling rate, seat height, pedal-crank length and the physical condition and determination of the cyclist. The traditional rowing motion, in which the rower sits still and brings into play only the muscles of his back, shoulders and arms, yields considerably less power than the cycling motion. The addition of the sliding seat, however, increases the power level of rowing to rival that of cycling—at least for periods of more than a few minutes. (The short-period advantage of cycling is lost after about a minute because of the limitations imposed by the human circulatory and respiratory systems.)

The inherent advantage of pedal-driven-propeller boats over rowed boats therefore lies chiefly in the fact that oars are a less efficient mechanism for channeling human energy to the propulsion of the boat. Furthermore, a rowed boat's uneven speed exacts more of a drag penalty than the smooth, continuous speed that can be achieved with a propeller.

Designers of fast human-powered watercraft have also attempted to minimize drag in novel ways. One way to virtually eliminate wave drag and at the same time reduce skin-friction drag is to submerge the hull; the operator would have to be supported above the water by narrow struts extending upward from the hull. The minimal-drag hull in this case is teardrop-shaped, with a length between three and four times its width.

Such a configuration is like that of a unicycle, and balancing would likewise be difficult, if not impossible, for the rider. Theodore Schmidt alleviated this problem somewhat by attaching four small outrigger hydrofoils to an experimental submerged-hull craft of his own design. A tricyclelike arrangement of three smaller submerged-buoyancy hulls would be stabler but not as efficient. Since the ratio of surface area to displacement gets smaller as displacement increases, one big hull has less surface area than three smaller hulls with the same total buoyancy.

The balancing problem of a single underwater hull could be solved by putting the operator in the hull—in effect creating a submarine. But a streamlined hull big enough to enclose a rider displaces much more water and has more surface area than a hull that provides just enough buoyancy to support a person's weight. Although it is not optimal for human-powered transport near the surface of the water, a human-powered submarine could be a great improvement over a skin diver with flippers. In the early 1950's a two-person human-powered submarine called the Mini-Sub, designed by Calvin Gongwer, was produced in limited quantities by the Aerojet-General Corporation. Pushed forward by twin 760-millimeter counterrotating propellers, the Mini-Sub reportedly could achieve speeds of seven knots—about three times the speed at which a diver can swim underwater.

Other designs seek to reduce the second major drag component, skin-friction drag, by employing dynamic lift to raise part of the boat out of the water and thereby reduce the boat's wetted surface area. Although dynamic lift does incur a drag penalty of its own, in many instances the reduction in skin-friction drag more than compensates for the drag generated as a by-product of the dynamic lift.

Human-powered water vehicles that take advantage of the dynamic lift achieved by planing are still in the imagination of designers, but another way to generate dynamic lift has been applied successfully: hydrofoils. Hydrofoils are underwater wings that produce lift in the same way as airplane wings produce lift. The required size of a hydrofoil wing is quite modest compared with airplane wings. For example, at a speed of nine knots something under a tenth of a square meter of foil area is needed to produce enough lift to support a single rider above the water. A hydrofoil designed to produce the same lift at twice the speed would require only a fourth as much area.

Although the small wetted area of hydrofoil wings results in minimal skin-friction drag, hydrofoils do incur a different type of drag. As the hydrofoil travels through the water it leaves behind a vortex wake, just as airplane wings do. The energy expended in generating the vortex wake is manifested as a drag called induced drag. Also, the spray kicked up by the vertical struts supporting the hydrofoil as they cut through the surface of the water results in additional drag.

Another major problem with hu-

HUMAN-POWERED WATERCRAFT display assorted shapes, construction materials and propulsion devices. Relatively primitive craft are poled (*a–c*) or paddled (*d, e*) and are made of such diverse natural materials as reeds, wood and animal skins. More modern craft are made of wood or metal and are rowed (*f*), an action that calls for the use of the arms, shoulders and back, or are paddled with foot pedals, bringing into play the strong leg muscles (*g*). Pedal-driven-propeller boats (*h–j*) have greater propulsive efficiency than either rowboats or paddle-wheel boats. Novel designs and materials have also cut back on the drag such craft encounter. The submerged hull of Theodore Schmidt's experimental craft (*j*), for instance, effectively eliminates the problem of wave drag.

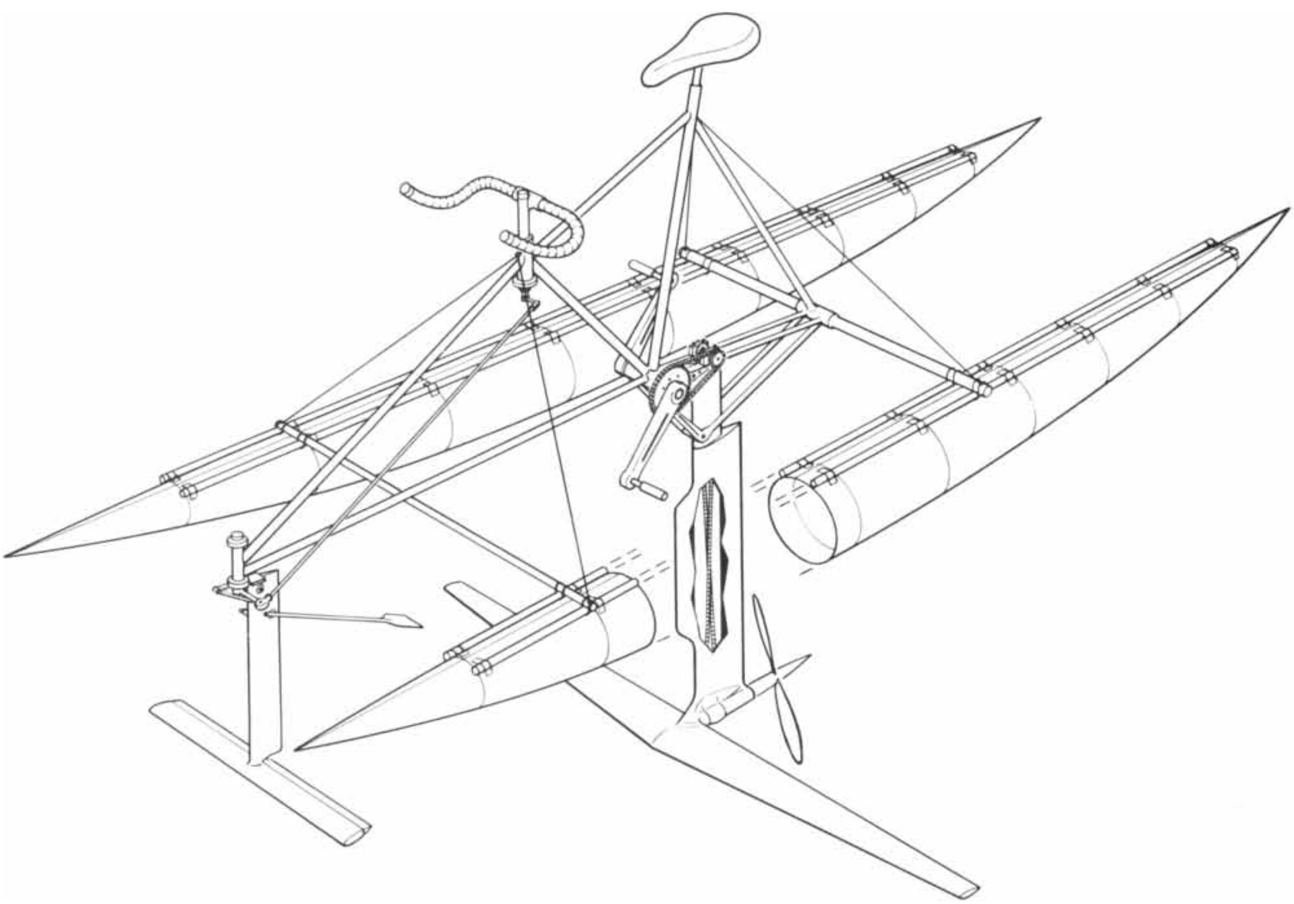
man-powered hydrofoil craft is that they need to reach relatively high water speeds before they can "take off," or lift themselves above the water. Since hydrofoils produce zero lift at zero speed, another support system, such as a displacement hull, is required for the initial and final phases of a "flight." A wing selected for all-out speed may have to be moving through the water at 10 knots before it can generate enough lift to support the craft and rider. This speed may well be impossible to achieve while the craft is still supported on the water by its displacement hull. A larger hydrofoil wing could reduce the takeoff speed, but the drag caused by the increased surface area would not allow the craft to go as fast.

Stacking hydrofoils so that smaller foils are placed below larger ones, as is done on motorized hydrofoil vessels, might circumvent the problem. A craft with such a tapered "ladder" arrangement of hydrofoils could take off at low speed on the large upper foils. Once sufficient speed has been attained so that the lift produced by the lower foil is enough to support the craft, the craft could rise up farther, raising the larger foil out of the water and thereby reducing drag. Because of the inherent difficulties associated with hydrofoils, human-powered hydrofoil craft do not have the same speed potential as human-powered airplanes, which have achieved speeds of more than 25 knots.

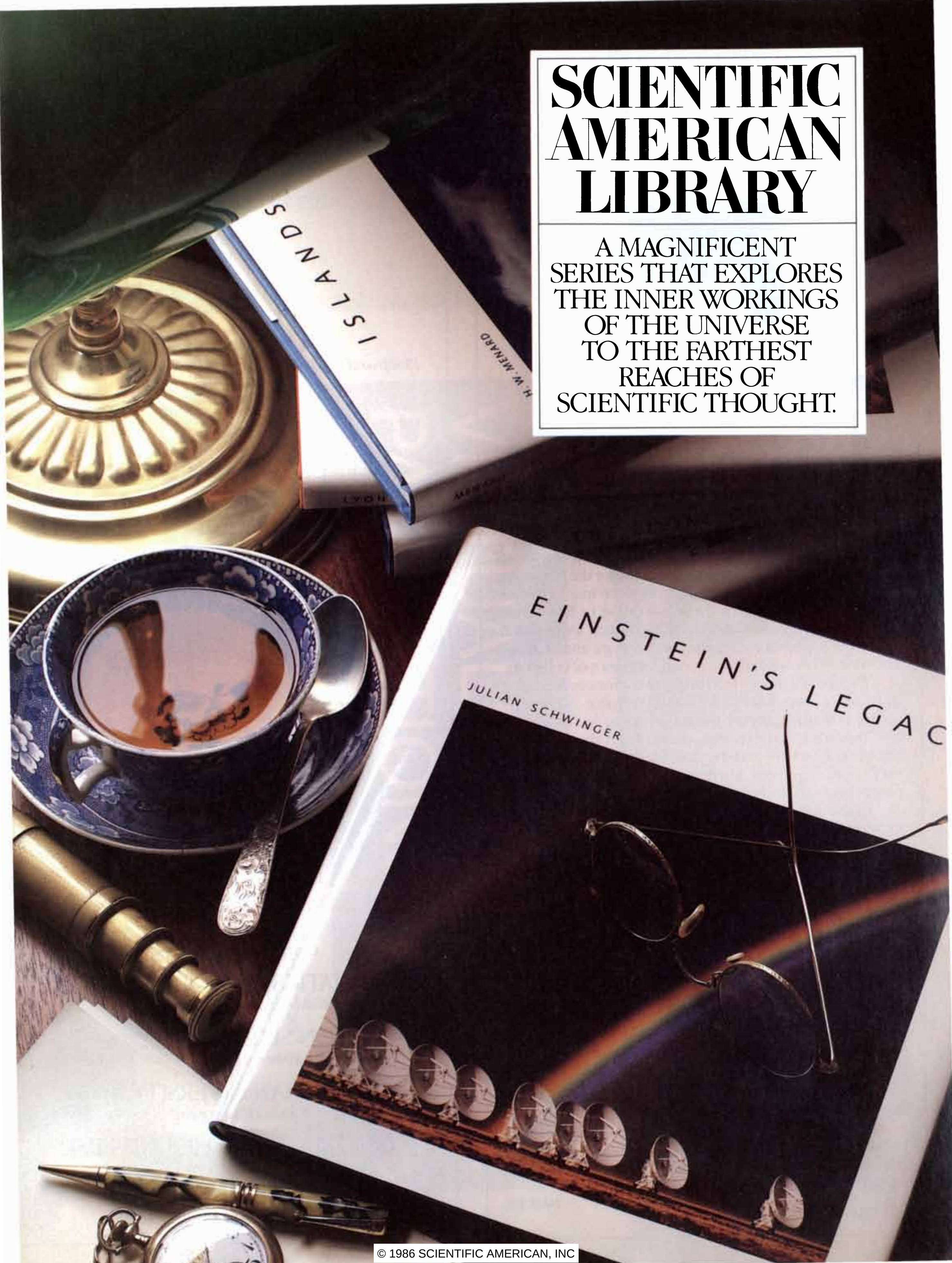
Until recently all human-powered water-speed records were held by displacement boats propelled by oars. With the intent of exceeding these speeds, two of us (Brooks and Abbott) in 1984 designed and built *Flying Fish I,* the first hydrofoil capable of sustained flight on human power alone. The sticky problem of initially getting the craft up to takeoff speed, which had plagued earlier designers, was bypassed initially by eliminating the need for a displacement hull. Flying speed was attained by catapulting the craft into the water from a floating ramp, much as jets are launched from aircraft carriers. Using this "flying start" launching method, cyclist Steve Hegg, an Olympic gold medalist, pedaled *Flying Fish I* a distance of 2,000-meters in six minutes 38 seconds, eclipsing the world record for a single rower by 11 seconds. The times, of course, are not directly comparable, because the rowing record was set from a standing start.

Flying Fish I has a high-efficiency, pedal-driven propeller and two slender wings supported by narrow vertical struts. The main wing, which carries 90 percent of the craft's weight, has a large wingspan (1.8 meters) to minimize induced drag and a small chord, or width, to reduce skin-friction drag. The smaller front wing has a configuration much like an inverted T and is lightly loaded; its main purpose is to provide stability and control. To this end it is fitted with a small, spatula-shaped device that automatically controls the depth of the wing. The device skates over the water surface, continuously adjusting a thin flap (analogous to the elevator on an airplane tail) to which it is linked.

The front wing strut doubles as a rudder and is connected to bicycle



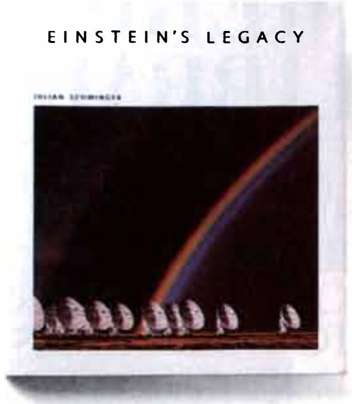
FLYING FISH II is a human-powered hydrofoil craft designed and built by two of the authors (Brooks and Abbott). The craft, powered by a pedal-driven, high-efficiency propeller, takes off at six knots and has a top speed of about 14 knots. It is ridden just like a bicycle. The first version of the craft did not have side pontoons and required a catapult-launch ramp to bring it up to takeoff speed. The pontoons on the current version support the craft so that it can now reach takeoff speed from a standstill. The depth at which the hydrofoils "fly" is controlled automatically by a spatula-shaped surface follower linked to a thin flap on the front hydrofoil. The craft has completed a 2,000-meter course approximately 10 seconds faster than the record for a single-rower racing shell.


SCIENTIFIC AMERICAN LIBRARY
A MAGNIFICENT SERIES THAT EXPLORES THE INNER WORKINGS OF THE UNIVERSE TO THE FARTHEST REACHES OF SCIENTIFIC THOUGHT.
ISLANDS
H. W. MENARD
EINSTEIN'S LEGAC
JULIAN SCHWINGER

EINSTEIN'S LEGACY

Julian Schwinger

Your Premier Volume



Nobel laureate Julian Schwinger tells the story of one of the twentieth century's greatest achievements—the theory of relativity—which is largely the work of one man: Albert Einstein.

The groundwork was laid in the seventeenth century by Isaac Newton, who unified motion on Earth and in the heavens in a framework of absolute space and time; then, in the nineteenth century by James Clerk Maxwell, who unified electricity, magnetism, and light. But it was left to the sixteen-year-old Einstein to glimpse, for the first time, that the theories of these two giants were incompatible.

Schwinger makes a lively narrative of Einstein's quest for the reconciliation of this conflict, a quest that led to the unification of matter and energy, and of space and time in his special and general theories of relativity.

The special theory has had its awesome confirmation in mankind's command of the nuclear force. Schwinger shows how the general theory, in turn, has stood up for 70 years to experiments drawn from the theory by Einstein, himself, and by his successors, employing ever more ingenious instrumentation and carrying the proofs ever farther beyond the decimal point.

In the last chapter, Schwinger describes space-age experiments made possible by the technologies that incorporate the theory itself. Their outcomes may not only secure further confirmation of Einstein's legacy, but face it with difficulties that lead on towards a still more comprehensive theory.

Julian Schwinger was awarded the Einstein Prize in 1951, the National Medal of Science in 1964, and the Nobel Prize for physics in 1965.

He is currently University Professor of the University of California, Los Angeles. He received his Ph.D. from Columbia University and has been on the faculty at Purdue University and Harvard University. Through the years, he has done theoretical work in various areas of both classical and quantum physics.

"I am happy to report, that his book makes for delightful and instructive reading.

"It is particularly gratifying that the subject's intricacies are so well conveyed here in simple language. Altogether, this well printed and pleasingly illustrated book is an ideal gift for the curious non-expert."

—NATURE

250 PAGES, 200 ILLUSTRATIONS

"The difference [between this and other book series] comes in the quality of the authors, who combine eminence in their fields with proven gifts of exposition. With the handsome volumes in front of us what response can there be but immediate surrender to this firepower?"

—Nigel Calder, NATURE

OTHER VOLUMES IN THE LIBRARY:

ISLANDS
H.W. Menard

SUN AND EARTH
Herbert Friedman

POWERS OF TEN
Philip and Phylis Morrison and the Office of Charles and Ray Eames

PERCEPTION
Irvin Rock

DRUGS AND THE BRAIN
Solomon H. Snyder, M.D.

THE SCIENCE OF MUSICAL SOUND
John R. Pierce

THE TIMING OF BIOLOGICAL CLOCKS
Arthur T. Winfree (in preparation)

SUBATOMIC PARTICLES
Steven Weinberg

THE SECOND LAW
P.W. Atkins

EXTINCTION
Steven M. Stanley (in preparation)

FIRE
John W. Lyons

THE LIVING CELL
Christian de Duve

FOSSILS
George Gaylord Simpson

MATHEMATICS AND OPTIMAL FORM
Stefan Hildebrandt and Anthony J. Tromba

ON SIZE AND LIFE
Thomas A. McMahon and John Tyler Bonner

HUMAN DIVERSITY
Richard Lewontin

CONSTRUCTING THE UNIVERSE
David Layzer

THE SOLAR SYSTEM
Roman Smoluchowski

handlebars for steering. The craft is ridden much as one would ride a racing bicycle. The structure that is normally above the water is, in fact, a modified bicycle frame.

Flying Fish II was developed as a refinement of the first version of our craft. We attached lightweight pontoon floats to it in the hope that an unassisted takeoff could be made from a standstill. This proved to be possible, and with practice acceleration from a standing start to the fully foil-borne mode took only three seconds. The craft also became much more practical because it could now "land" on, as well as take off from, its floats. (The catapult-launched *Flying Fish I* gave the rider a dunking whenever he stopped pedaling.)

Aboard the *Flying Fish II* one of us (Abbott) recorded a time of six minutes 39.44 seconds over a 2,000-meter course from a standing start—about 10 seconds faster than the single-person rowing-shell record. From a flying start the hydrofoil watercraft also was able to sprint 250 meters in 38.46 seconds, reaching a maximum speed of approximately 13 knots.

The time is ripe for a technological revolution in human-powered recreational watercraft. Laser International has just introduced the Mallard, a partially enclosed, seaworthy boat designed by Garry Hoyt. Several new pedaled catamarans and proas (boats that have one main hull and a smaller stabilizing outrigger) offer rough-water seaworthiness and impressive speed. Jon Knapp of Saber Craft has designed and built a propeller-driven proa that is faster than a shell in rough water but, unlike a shell, requires no special skills to operate. The Dorycycle, a propeller-driven single-hulled craft designed by Philip Thiel, provides good load-carrying capacity at speeds twice that of the rowed dory from which it was derived.

Whether or not hydrofoil craft become popular for recreation, there seems to be little doubt that they will figure prominently in the next round of breaking records. The International Human Powered Vehicle Association encourages competition in human-powered vehicles—on land, on sea and in the air—without any arbitrary limits placed on their design. Such competition will push the speed of human-powered hydrofoil craft ever higher. It is not farfetched to envision such craft reaching speeds as high as 20 knots—one and a half times as fast as the speeds attained by *Flying Fish II* and significantly faster than the speeds attained by racing shells powered by eight athletic oarsmen.

A Roman Apartment Complex

In the second-century Garden Houses of Ostia a rigorous geometry prevailed. From the overall plan to individual mosaics, a particular pattern, rich in philosophical suggestiveness, underlay the design

by Donald J. Watts and Carol Martin Watts

Near the mouth of the Tiber River, under the flight path of Leonardo da Vinci airport, stand the excavated brick remains of the ancient Roman port of Ostia. Among the most fascinating ruins at this rich archaeological site are those of a neighborhood of apartments, shops and gardens called the Case a Giardino, or Garden Houses. The Garden Houses are impressive in part because they were a planned complex; they were surprisingly similar to a modern development. Even in their ruined state, with walls standing less than one story high and weeds growing in the spaces that at one time must have been beautiful gardens, they convey a palpable sense of order and design.

As part of a larger study of ancient Roman houses we analyzed the Garden Houses in detail. In the process we found the key to their design: a single geometric pattern that recurs at all scales from the overall configuration of the buildings to the layout of floor mosaics. The pattern, which is based on the square and a particular way of dividing it, is called the sacred cut. By ensuring proportional relations among the parts of the complex and of the parts to the whole, the sacred cut lends unity and harmony to the design. To the ancient Romans, and in particular to the unknown architect of the Ostia Garden Houses, the pattern may also have had a more profound philosophical significance.

In its heyday during the first and second centuries A.D. Ostia was a densely populated city, its riverfront busy with galley traffic, its warehouses stocked with wheat, oil and wine destined for the imperial capital. Silt, along with the decline of the Roman empire, killed Ostia. The Tiber delta gradually became unnavigable, the coast advanced westward and the town was left behind. By the ninth century it had been abandoned; over the next millennium periodic floods buried its ruins in mud.

The slow silting up of the Tiber created at Ostia what the catastrophic eruption of Vesuvius in A.D. 79 created at Pompeii and Herculaneum: a treasure of information on the domestic architecture of ancient Rome. (Continual rebuilding has destroyed most of the ancient houses in the imperial city itself.) The different sites are complementary. In Pompeii and Herculaneum there are many well-preserved examples of the traditional single-family house, the *domus,* with its characteristic arrangement of rooms around a central atrium. In densely populated Ostia, on the other hand, as in Rome itself, only the very wealthy could afford a *domus;* the middle and lower classes lived in three-to-six-story apartment buildings called *insulae.* The excavation of Ostia in the first part of this century uncovered many *insulae,* of which the Garden Houses are the most impressive examples.

Although it is not known who designed the Garden Houses, it is known when they were built: in about A.D. 128, during the reign of the Emperor Hadrian. The construction of a new artificial harbor near Ostia by Hadrian's predecessor Trajan had increased the city's commercial importance as the gateway for the import of goods to Rome. To accommodate the immigrants who flocked to Ostia (its population grew to about 50,000) more than half of the city was rebuilt; *domus* were replaced by massive *insulae* of vaulted, brick-faced concrete. Trajan and Hadrian were the patrons for much of the building boom, but the Garden Houses, the largest single project, were probably a private investment. They appear to have been intended for well-to-do merchants: they were desirably situated near the sea and away from the busy riverfront, and the apartments were unusually large and well decorated, with a few of the units being particularly luxurious.

The Garden Houses complex consisted of a continuous building perimeter, whose irregular shape was probably dictated by preexisting streets, surrounding a rectangular courtyard approximately 100 meters long by 80 meters wide [*see illustration on page 135*]. Within the courtyard were two prominent, freestanding buildings. Each consisted of standardized apartments, with four apartments on the ground floor. Each apartment had a central space and main path of circulation that took the place of an atrium and that may have been called a *medianum.* At the ends of this long, light-filled room, which faced the exterior wall, there were two other major rooms. Several smaller, windowless bedrooms adjoined the *medianum,* from which they got both light and air.

The buildings on the perimeter contained at least nine additional ground-floor apartments as well as about 40 shops. All the ground-floor apartments in the courtyard buildings and most of those in the perimeter buildings had an internal stairway leading to a mezzanine. Separate stairways opening directly onto the courtyard or the street gave access to apartments in the upper stories. Little is known about the upper stories; only the ground floors of the Garden Houses have sur-

GARDEN HOUSES were excavated along with the rest of Ostia in the first part of this century. The apartment complex, which was built in about A.D. 128, consisted of a continuous building perimeter surrounding two buildings in a rectangular courtyard. The photograph was made from above the main east-west axis of the complex; the front (east) gate is at the bottom. The "House of the Muses" in the northeast corner has been restored.

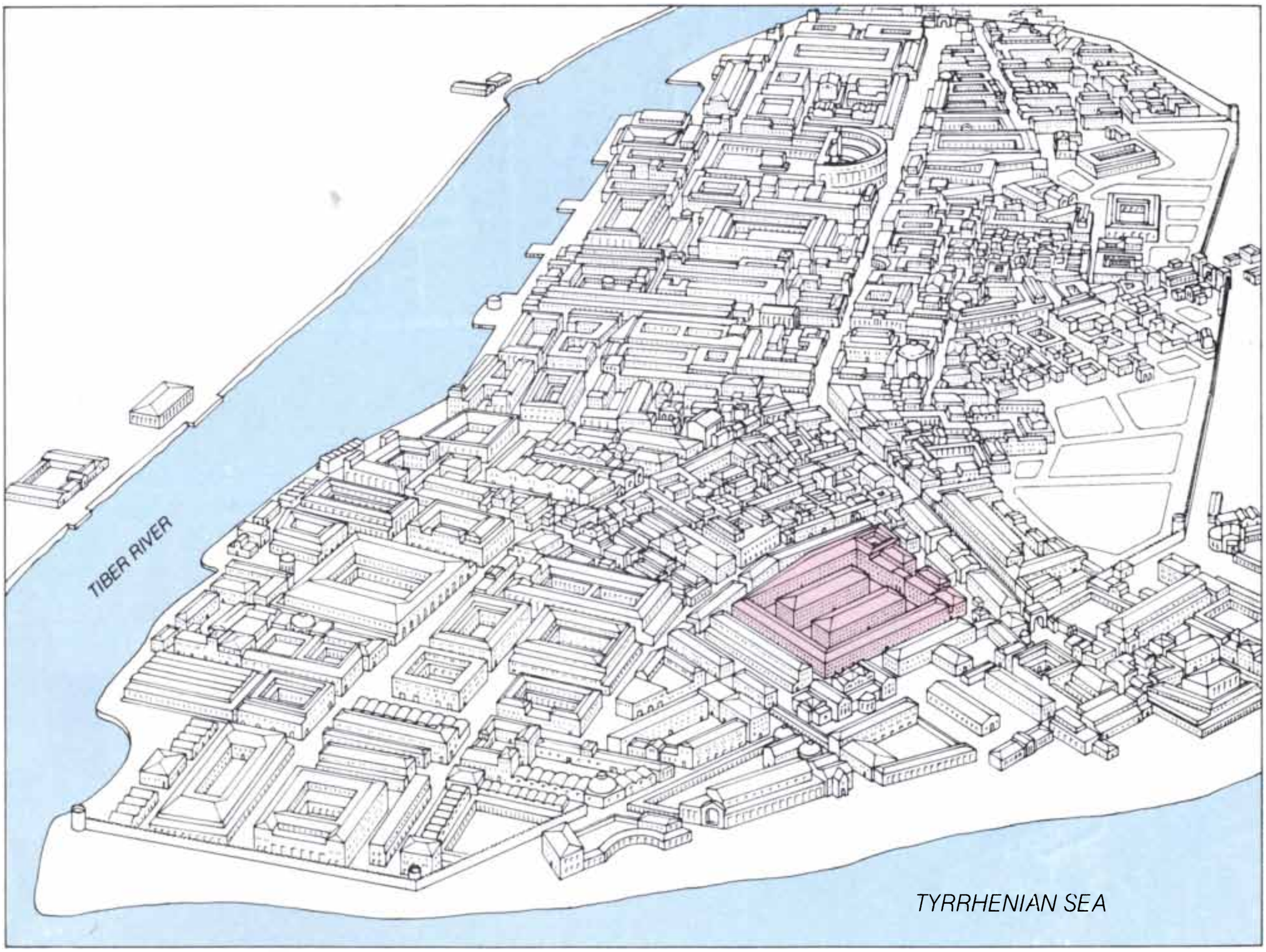
vived. The thickness of the walls indicates the buildings were probably four stories high, including mezzanines. Depending on the size of the upper apartments and on whether they too had mezzanines, the complex could have had anywhere from 40 to 100 apartments. It probably housed between 400 and 700 people, including shopkeepers who lived in their shops.

The complex is oriented approximately to the compass points; its front and the main gate face east. Given the irregular shape of the perimeter, it is the large rectangular courtyard that imposes geometric order on the complex. Its major axis follows the corridor between the two courtyard buildings and connects the large east gate to a smaller gate on the west side. The minor north-south axis is defined by covered passageways bisecting the courtyard buildings; its lesser importance is indicated by the fact that the north and south gates are offset from the axis. The two axes cross at the center of the complex. It is likely that the center was further emphasized by the landscape plan—a focus on the center is a hallmark of Roman design—but one cannot say for sure: all that remains of the original gardens are six fountains, near the east and west edges of the courtyard, from which the residents drew their water.

Both the fountains and the courtyard buildings turn out to be key elements in the geometric ordering of the Garden Houses. As we measured and inspected all parts of the complex we became increasingly aware of the architect's hand, of a deliberateness in the size and positioning of the courtyard buildings. In attempting to account for this sense of order we considered and rejected dozens of geometric schemes on which the architect might conceivably have relied. Finally we hit on the idea of the sacred cut.

The sacred cut is a simple geometric operation [*see illustration on page 136*]. It can be carried out with a straightedge and a compass, both of which are known from the writings of the architect Vitruvius to have been employed by Roman builders in determining ground plans. One begins by drawing a square (the reference square) and its diagonals. Next one draws quarter circles centered on the corners of the square, each with a radius equal to half of the diagonal. The arcs pass through the center of the square and intersect two adjoining sides; together they cut the sides into three segments, with the central segment being larger than the other two. By connecting the intersection points one can divide the reference square into a nine-part grid. At the center of the grid is another square (the sacred-cut square) that can serve as the foundation for the next sacred cut.

The term sacred cut is not an ancient one; it was coined some 20 years ago by the Danish scholar Tons Brunés. As it happens, the length of an arc in the



OSTIA was a boom town in the second century A.D. The construction of a new artificial harbor (not shown) on the Tyrrhenian Sea had increased Ostia's importance as the major port of Rome; goods imported from around the Mediterranean were shipped from the harbor to Rome, some 25 kilometers away, by way of the Tiber River. As Ostia's population swelled to about 50,000, many apartment buildings were erected, including the Garden Houses (*color*). The drawing is based on a model derived from the ruins.

sacred cut is equal, to within about .6 percent, to the length of the diagonal of the rectangle that is half of the reference square. According to Brunés, this near-equality of an arc and a straight line may have convinced ancient builders they had found an empirical way of squaring the circle, that is, of constructing a square with the same perimeter as a given circle, or vice versa. (A circle cannot be squared precisely because the perimeter of a square is a rational number, whereas the circumference of a circle is proportional to the irrational number π.) To ancient geometers the circle symbolized the unknowable, spiritual part of the universe and the square represented the comprehensible world. Squaring the circle was a means of expressing the unknowable through the knowable, the sacred through the familiar. Hence the term sacred cut.

Although Vitruvius does not discuss the sacred cut specifically, he does mention the importance of geometric patterns as a way of achieving what he considers the most important quality of a good design: proportional relations among the various elements. Brunés has traced the application of the sacred cut from the ancient Egyptians through the Greeks and Romans to medieval Europe. (He believes the idea was transmitted from Egypt to Greece in the sixth century B.C. by the philosopher Pythagoras.) Most notably, he has found the sacred cut in the design of the Roman Pantheon, which was built at about the same time as the Garden Houses of Ostia.

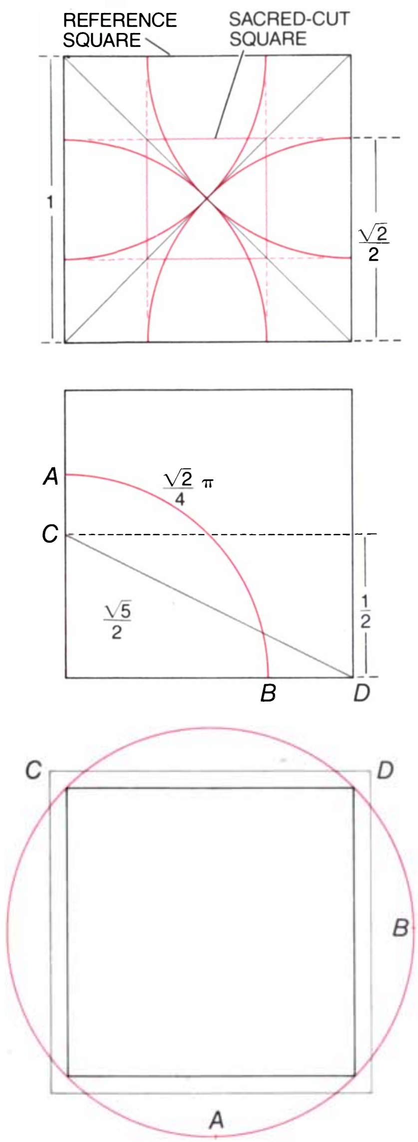
Sacred cuts are pervasive in the design of the Garden Houses, beginning with the overall layout of the complex. A sequence of three cuts of three reference squares define both the size and the position of the courtyard buildings and thereby also of the open spaces in the courtyard [*see illustration on page 137*]. In the process the cuts emphasize the center and the axes of the apartment complex, giving it a formal unity.

The largest reference square follows the perimeter of the complex. It does so only approximately because the shape of the perimeter is irregular. That the square has not been chosen arbitrarily, however, can be seen from the fact that a circle drawn within the square and tangent to its sides just touches the corners of the courtyard. The center of both the circle and the square marks the center of the complex. The sacred cuts of the east and west sides of the square establish guidelines that coincide with the northernmost and southernmost walls of the two courtyard buildings. These walls reaffirm the dominance of the east-west axis established by the shape of the courtyard.

Proceeding toward the center, the next reference square is based on the width of the courtyard. Because the courtyard is rectangular, the east and west sides of the square do not follow

GROUND-FLOOR PLAN of the Garden Houses complex shows its basic geometry. The major east-west axis and the minor north-south axis cross at the center of the complex. The buildings were probably four stories high. Each of the 17 ground-floor apartments consisted of several rooms surrounding a central space, which may have been called the *medianum* (*light color*). The complex also included many shops (*dark color*). Entrances to the gardens in the courtyard were at the sides and in the corners of the complex.

the edges of the courtyard; instead they coincide precisely with the inner edges of the six fountains. Sacred cuts of this square define the next constructional planes: the spines of the courtyard buildings, along which thick party walls separated back-to-back apartments. The party walls further emphasize the east-west orientation of the complex.



"SACRED CUTS" of a reference square are constructed by drawing arcs that are centered on the corners and pass through the center of the square (*top*). By connecting the points where the arcs cut the sides one obtains a nine-part grid; the central square is called the sacred-cut square. The length of each arc *AB* is equal, to within .6 percent, to the length of the diagonal *CD* of half of the reference square (*middle*). Hence the sacred cut provides an approximate method of squaring a circle: the perimeter of a square composed of four lines *CD* is nearly equal to that of a circle composed of four sacred-cut arcs (*bottom*).

The sacred-cut square at the center of the second reference square serves as the third and last reference square. Its sacred cuts in turn mark the position of the innermost walls of the courtyard buildings, the walls closest to the center of the complex, which bound the central east-west corridor. When the three reference squares are superposed, they nest concentrically inside one another, and their sacred cuts appear to unfold like the layers of an onion.

Sacred-cut geometry not only unifies the complex as a whole but also shapes the design of individual buildings. For example, the diagonal of the smallest sacred-cut square in the above sequence is equal to the outside width of the courtyard buildings. Furthermore, each building is five of these squares in length. The central square in the row of five encloses the stairways and the entrance halls of the building, whereas the squares on both sides define the living spaces.

The geometry extends further to individual apartments in the courtyard buildings [*see illustration on page 138*]. Here the sacred cut serves not as a method of positioning design elements but as a source of a series of proportional, whole-number dimensions that regulate the plan of the apartments. The reference square in this case has a diagonal of 58 Roman feet, which is the inside rather than the outside width of the courtyard buildings. (A Roman foot is .295 meter, or roughly 11½ inches.) The square is 41 Roman feet on a side. Two successive sacred cuts yield line segments of 17, 12, seven and five Roman feet.

These dimensions or multiples of them figure prominently in the apartment plan. The interior width of an apartment is 28 feet, or four times seven feet. The width of the *medianum* and of the bedrooms is half of that, or 14 feet. The windows in the largest room are five feet wide; the space between them is two feet, and so the width of a window unit is seven feet. Hallways inside the apartments are five feet across. The width of the public space, consisting of the covered passageway and the stairway to the upper stories, is 17 feet.

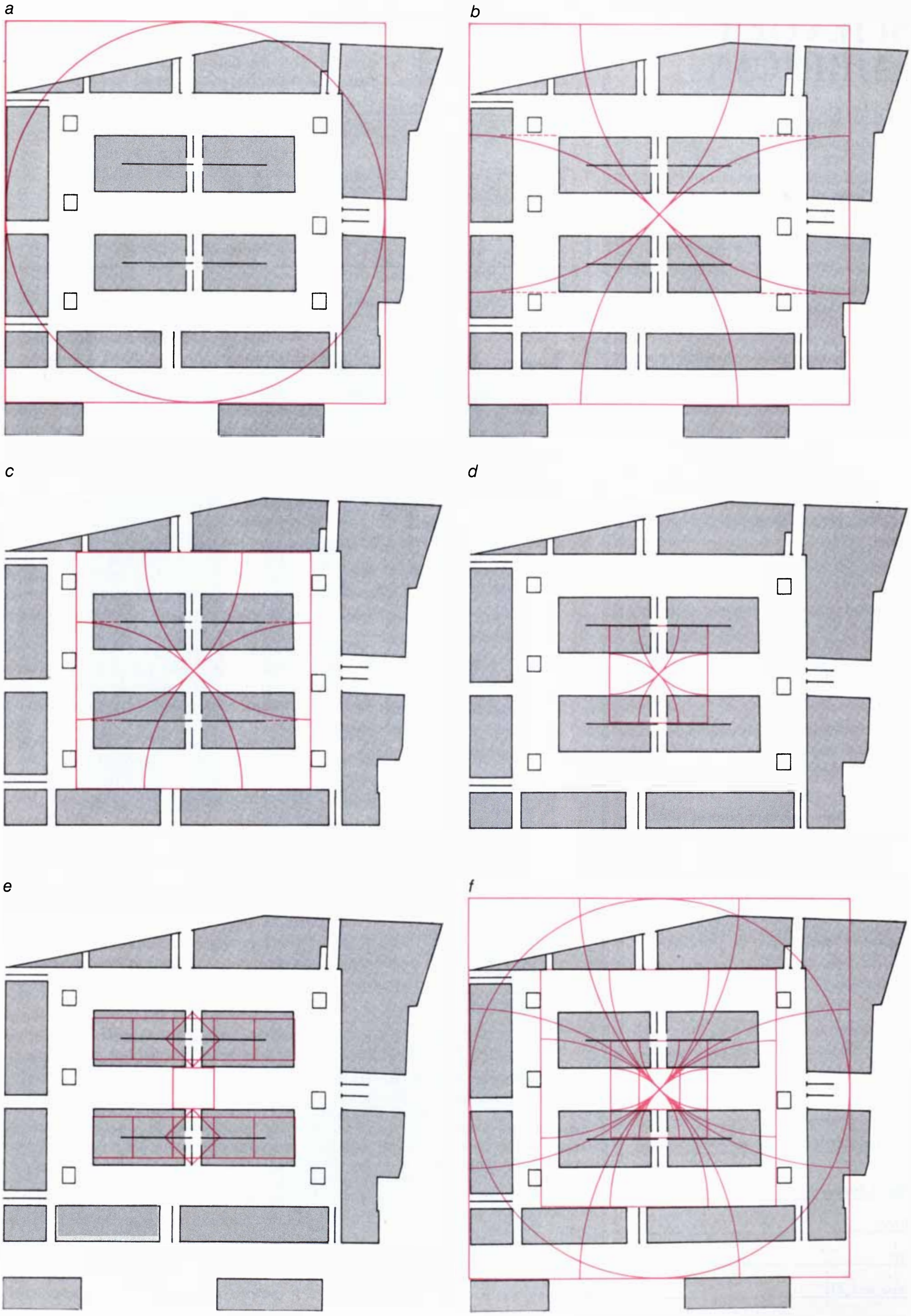
The dimensions must have held more than a merely practical significance for the architect of the Garden Houses: they are precisely the numbers generated by the Pythagorean procedure for approximating the irrational square root of 2. The procedure begins with a square one unit on a side. The diagonal of the square, which according to the Pythagorean theorem is equal to the square root of 2, is given the approximate value of 1. Next the diagonal is added to the side of the square to generate the side of a larger square; the diagonal of the second square is roughly equal to the diagonal of the first square added to twice the side of the first square. By constructing a series of squares in this manner and dividing the approximate diagonal of each square by its side, one obtains the series of ratios 1/1, 3/2, 7/5, 17/12, 41/29 and so on. The series converges toward the square root of 2.

The same series of ratios can be generated by squares and sacred cuts, because the ratio of the side of a square to the radius of its sacred-cut arc is also equal to the square root of 2. (In this case the side of each square in the series is constructed by adding the side of the previous square to twice its sacred cut.) The fifth square in the series has a side of 41 and a sacred cut of 29—as does the square underlying the apartment plan of the Garden Houses.

By basing the plan on a series of numbers that approximate the irrational square root of 2, the architect of the Garden Houses was making a philosophical statement akin to that of squaring the circle (which is equivalent to approximating π). In both instances the sacred cut is the means of expressing the irrational and undefinable by the rational and definable.

At the smallest scale in the Garden Houses the sacred cut underlies the design of individual floor, wall and ceiling decorations. None of the decorations in the courtyard buildings are well preserved, but some have survived in the perimeter. In a building called the House of the Muses, at the northeast corner of the complex, a

GEOMETRIC ORDER of Ostia's Garden Houses complex is established by three successive sacred cuts. A square roughly congruent with the perimeter of the complex encloses a circle that touches the corners of the courtyard (*a*). Sacred cuts of the east and west sides of this reference square determine the position of the outer walls of the courtyard buildings (*b*). The second reference square, concentric with the first, is defined by the width of the courtyard and the positions of the fountains; the sacred cuts of its east and west sides guide the placement of the party walls along the spines of the courtyard buildings (*c*). The third reference square is the sacred-cut square of the second, and its cuts define the innermost walls of the courtyard buildings (*d*). The buildings are precisely five times as long as the final sacred-cut square, and their width is equal to its diagonal (*e*). A superposition of all the sacred cuts shows how they unfold from a common center, thereby emphasizing the major east-west axis of the complex (*f*).


a
b
c
d
e
f

SCIENTIFIC AMERICAN

CORRESPONDENCE

Offprints of more than 1,000 selected articles from earlier issues of this magazine, listed in an annual catalogue, are available at $1.25 each. Correspondence, orders and requests for the catalogue should be addressed to W. H. Freeman and Company, 4419 West 1980 South, Salt Lake City, UT 84104. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.25 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970. Specify CCC Reference Number ISSN 0036-8733/86. $1.25 + 0.00.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, NY 10017. Manuscripts are submitted at the authors' risk and will not be returned unless they are accompanied by postage.

Advertising correspondence should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, NY 10017.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name ______________________

New Address

Street ______________________

City ______________________

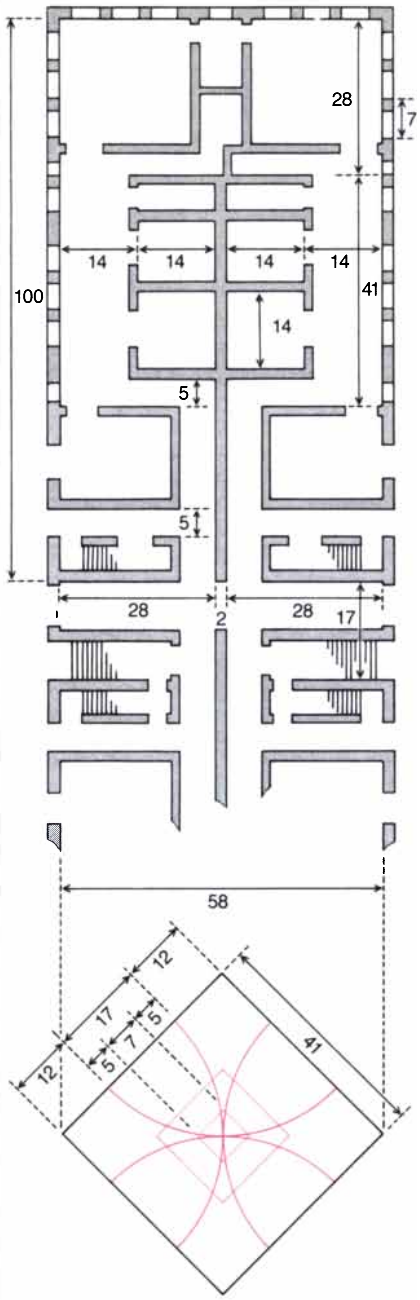
State and ZIP ______________________

Old Address

Street ______________________

City ______________________

State and ZIP ______________________



X	Y	Y/X ($\sqrt{2}$ = 1.4142. . .)
1	1	1
2	3	1.5
5	7	1.4
12	17	1.4166. . .
29	41	1.4137. . .

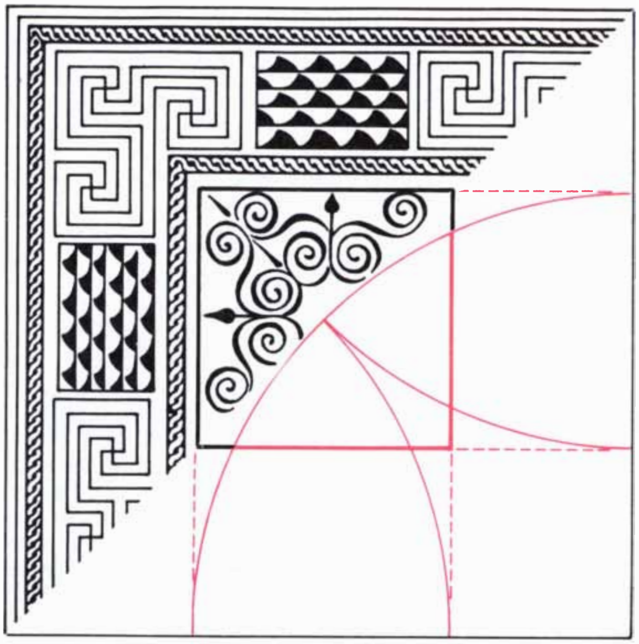
APARTMENT PLANS in the courtyard buildings are regulated by the sacred cuts of a square whose sides are 41 Roman feet long and whose diagonal is equal to the interior width of the buildings (58 feet). Two back-to-back apartments, separated by a two-foot party wall, are shown. The dimensions established by successive sacred cuts of the square appear throughout the apartments; the number 7 and its multiples are particularly frequent. The number series generated by the sacred cuts is the same as the one used by ancient geometers to approximate the irrational square root of 2.

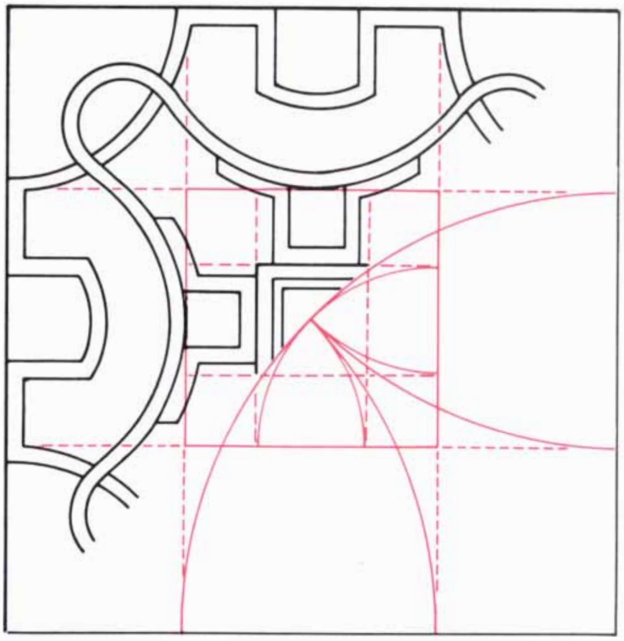
square floor mosaic fills most of one room. In the center of the mosaic there is a square medallion; its size is defined by the sacred cut of the outer border. Similar compositions are found in other parts of the Garden Houses. In some cases the mosaics in different rooms of the same apartment seem to be related in size and geometry. For example, the mosaic in one room may be the size of the sacred-cut square of the mosaic in a larger room.

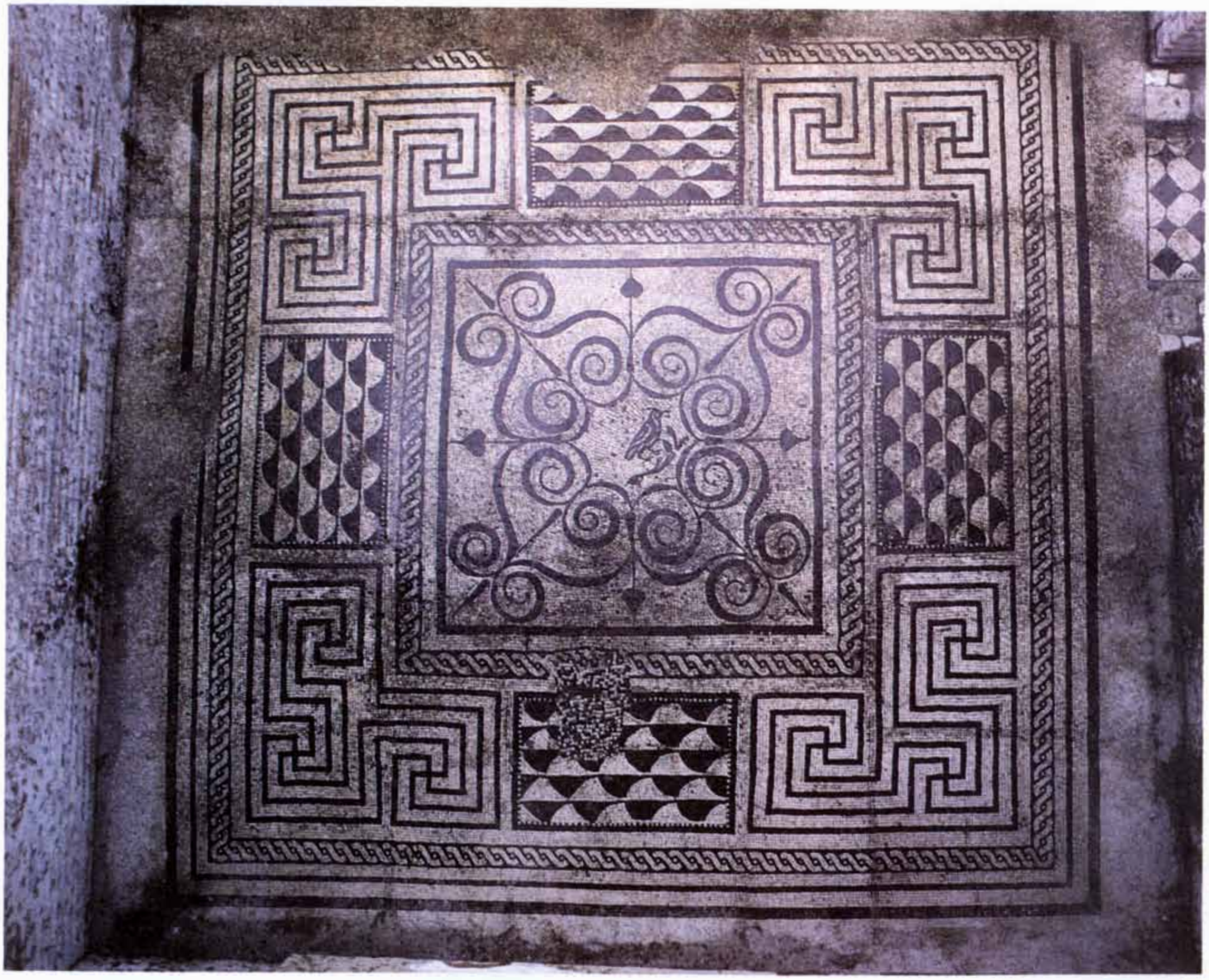
In addition to having floor mosaics it is likely that most of the rooms in the Garden Houses were decorated with brightly colored ceiling and wall paintings. None of the ceiling paintings are intact, but there are enough fragments of one of them to make it possible to reconstruct its composition. Like the floor mosaics it is square and centralized. Sacred cuts of the outer square determine the positions of two sets of arcs; the sacred-cut square of the outer border is then cut again to define a central medallion.

Paintings on the interior walls also give evidence of the sacred cut. To be sure, many of the walls are not square, being either wider than they are high or higher than they are wide. Nevertheless, in almost all the preserved paintings the composition appears to be based on a conceptual reference square whose sides are the width of the wall. The geometry of the sacred cut not only determines the placement of major elements of the composition but also, as in the apartment plans, provides a set of numbers that govern crucial dimensions in the painting.

The sacred cut is only one of many geometric patterns, most of them strongly centralized and axial, that are found in Roman design. Yet its insistent repetition at all scales in the Garden Houses of Ostia suggests that it had a kind of ritualistic importance. We are now looking for the sacred cut elsewhere in Roman architecture. Already our measurements have revealed its presence in the plan and decoration of many single-family houses in Pompeii and Herculaneum, which were built between one and three centuries before the Garden Houses; apparently the sacred cut was a feature of the *domus* that was retained when Roman architects began designing high-density housing. And of course the application of the pattern went beyond domestic design. As we mentioned above, Brunés has discovered it in such public buildings as the Pantheon. It may even have been used in urban planning: we have found tentative evidence that an entire Roman city, built in the Middle East in the first century A.D., was laid out using a geometry determined by the sacred cut.







MOSAICS AND PAINTINGS in the Garden Houses are in many cases laid out according to the geometry of the sacred cut. The photograph shows a floor mosaic in the House of the Muses. The medallion at the center of the mosaic is the sacred-cut square of the outer border (*top left*). In the design of a ceiling painting (*top right*) sacred cuts of the outer border determine the radius of the corner arcs, and the large side arcs are tangent to the sacred-cut square. The cuts of that square frame the central medallion.

THE AMATEUR SCIENTIST

Methods for going through a maze without becoming lost or confused

by Jearl Walker

What is the best way to penetrate a maze from an entrance to an interior goal? Is there any way to reach the goal and return to the entrance without traveling any path twice? Can you avoid endlessly circling within a maze? Suppose you realize you are lost. How can you find your way back to the entrance without wandering hopelessly farther into the maze? In examining these questions I shall introduce some delightful color mazes developed by Minotaur Designs in England.

Several terms require definition. An entrance to a maze is where you begin your search; it is usually on the perimeter of the maze. The goal is the point for which you search. It can be anywhere in the maze; it may be an exit. A node is an entrance, a goal or any point where a path branches or dead-ends. The path between successive nodes is called a branch. A route is a sequence of branches. A wall is the side of a path. In a cave system the wall is obvious. In a garden maze it may be a hedge or shallow mounds that border and define the paths of the maze.

Some mazes have nodes only at the entrance and the goal. You follow a sinuous route to the end without any risk of becoming lost. Mazes with additional nodes are harder to solve because they require decisions about which branches to explore. If you choose branches at random, you may end up circling aimlessly, never reaching the goal or even regaining the entrance. Some mazes limit your explorations. For example, you may be allowed to travel along a branch only once or only in a certain direction. You may also be required to visit certain places within the maze in a special order. If the maze has several possible routes to the goal, you may be required to find the route that involves the lowest number of nodes. I call this technique a minimal route.

If you have a map of a maze, you can always find a direct route from the entrance to the goal by trial. The task is easier if you shade the dead ends. As you extend the shading a direct route becomes apparent.

What if you enter a maze without a map or even the means of making one? How should you make decisions at the nodes to avoid becoming lost?

One technique is to choose the same direction at each node. For example, you could decide always to choose the branch farthest to the right. If it dead-ends, you return to the node and choose the branch next farthest to the right. You may end up traveling through branches twice, once in each direction, but eventually you will reach the goal. On your way back to the entrance you could either continue to choose the rightmost branch at each node, possibly traveling through new regions of the maze, or you could always choose the leftmost branch and thus exactly retrace your route to the goal. I call the technique of consistently choosing branches on either the right or the left the hand rule.

The hand rule works only for mazes that are said to be simply connected. The term indicates that the maze does not contain a closed route: a route that loops back on itself. A closed route is created by a walled island that does not connect with other walls in the maze. A maze with one island or more is said to be multiply connected.

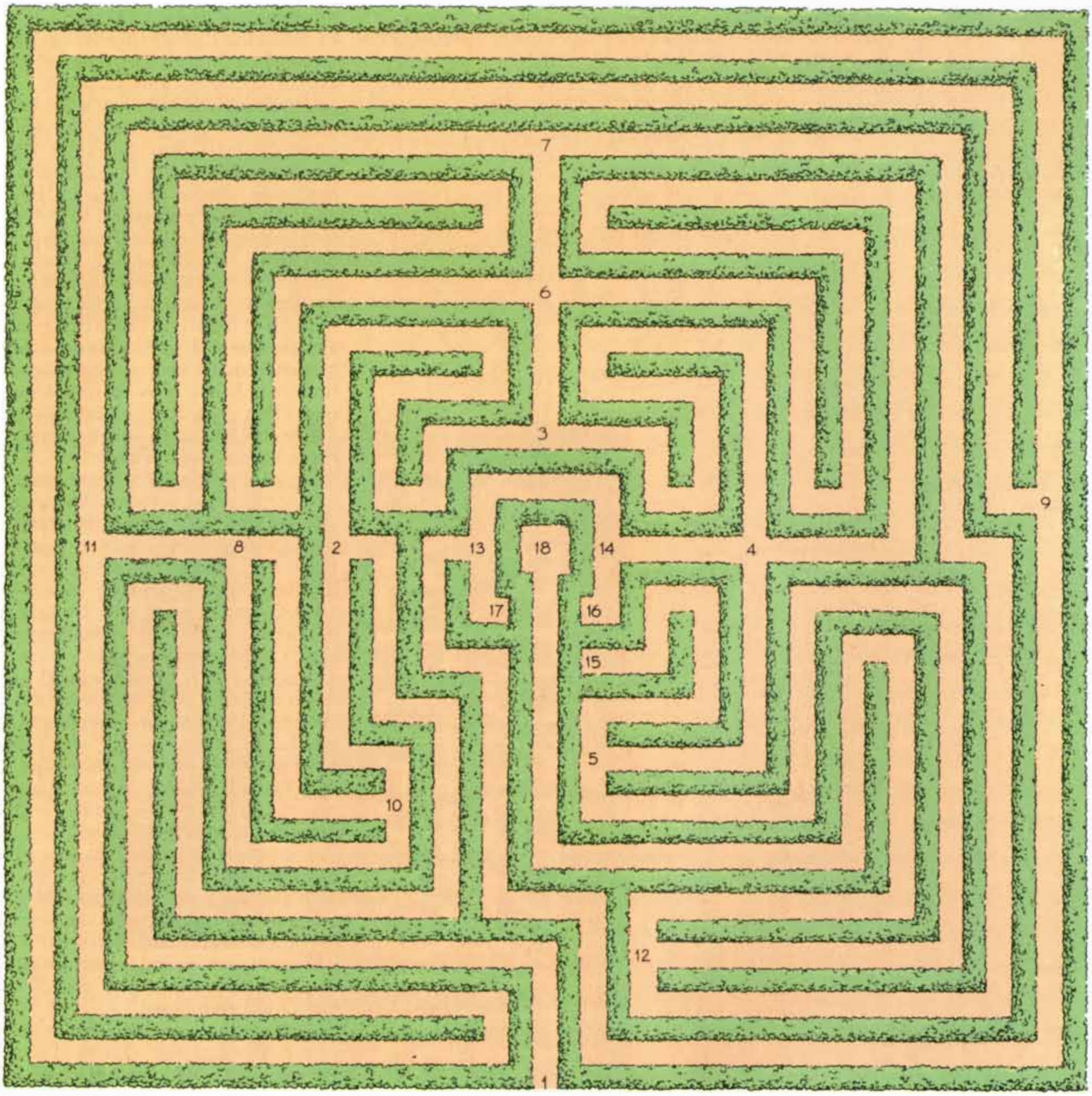
The first multiply connected hedge maze was constructed in the 1820's at Chevening in Kent, England. It has eight interlocking islands [*see top illustration on opposite page*]. The nodes are numbered, with node 1 at the entrance and node 18 at the goal. Suppose you enter this maze without benefit of a map and employ a right-hand rule at each node. You travel through a sequence of nodes 1–2–3–4–14–13–9–11–8–10–2–1, never reaching the goal. You may even think you have seen all of the interior, which would indeed be the case if you returned to the entrance in a simply connected maze.

A hand rule for exploring a multiply connected maze will fail only if the maze has a closed route around its entrance or goal. All other closed routes pose no problem. Suppose you approach an interior island [*see illustration at bottom left on opposite page*]. If you consistently employ either the left-hand or the right-hand rule, you cannot be trapped in a closed route around the island. In order to be trapped going clockwise you must err by first choosing the leftward branch at node *a* and then choosing the rightward branches thereafter. To become trapped going counterclockwise you must err by first choosing the rightward branch at node *a* and then choosing the leftward branches thereafter. (This principle will be of little comfort when you enter a maze without knowing whether you are already traveling around an island or whether the goal is protected by an island.)

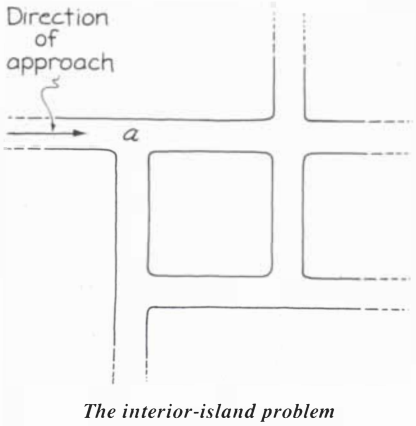
The complexity of a maze can be greatly reduced if a map of it is topologically distorted into a simpler pattern called a network. In such a pattern all the nodes and interconnections are retained but the twisting of pathways is eliminated. In a network for the Chevening maze a direct route of 1–2–3–4–5–12–18 to the goal is easy to see. Another route, 1–2–3–6–7–12–18, is just as good because it involves the same number of nodes. Each loop in the network is produced by an island.

In a network a direct route from the entrance node to a goal is laid out in a straight line. A dead-end branch leaves the direct route and does not return. An island at the entrance creates a route that encircles the entrance node. It can intersect the direct route at one node with four branches or at two nodes. The hand rule for exploration fails with this closed route. Another loop encircles the goal, again nullifying the hand rule. An encircling route can also be drawn as passing under the direct route. Interior islands create loops that intersect the direct route at either one node or two nodes. Note that if you pass into such a loop with the hand rule, you eventually escape from the loop and continue along the direct route toward the goal. More complicated loops that make additional connections with the direct route can also be escaped from.

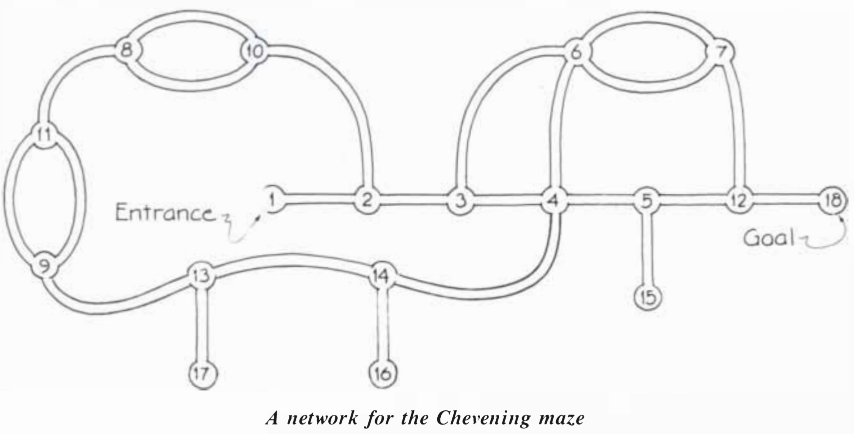
Since the hand rule does not guarantee success in reaching the goal, how should one explore a maze? Several techniques are known, but the one most widely cited is credited to one



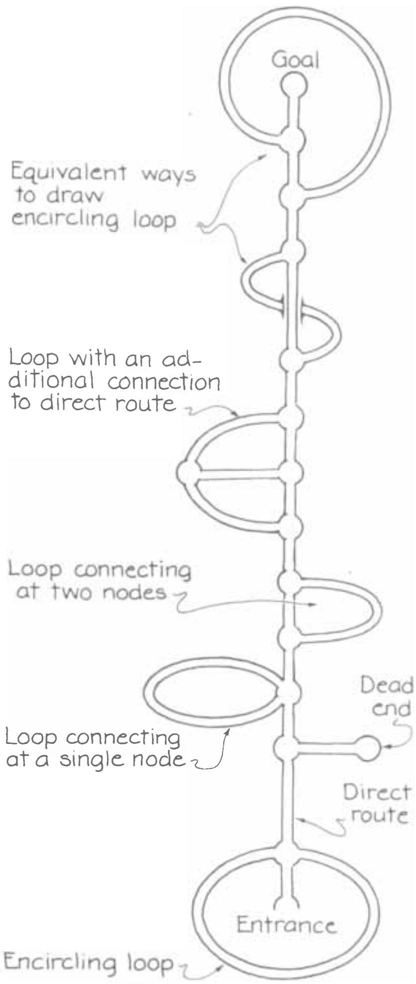
A map of the hedge maze at Chevening



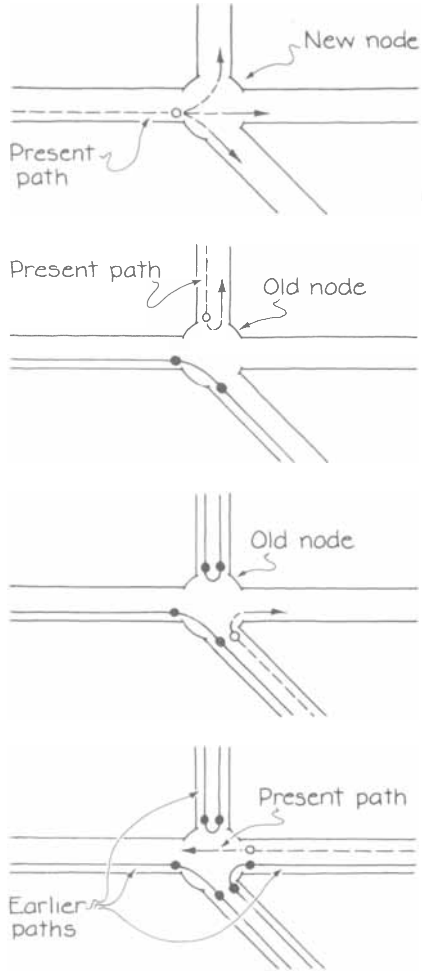
The interior-island problem



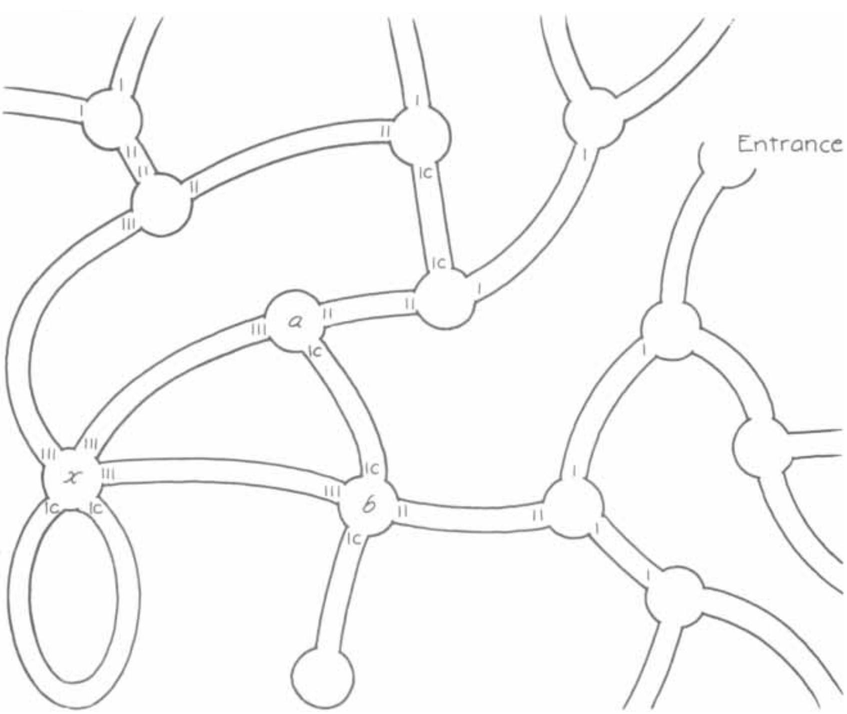
A network for the Chevening maze



The basic elements of a network



Trémaux's rules for exploring a maze



Numbering to escape a maze

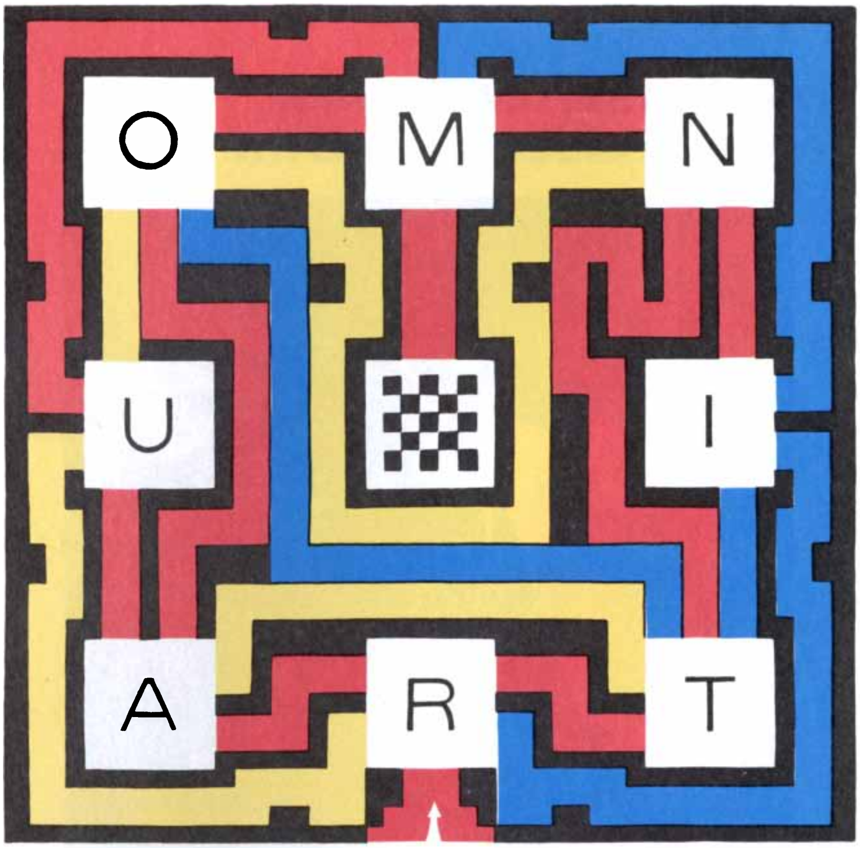
M. Trémaux in the 1882 publication of Édouard Lucas's *Récréations mathématiques.* In the illustration of it at the left the terms old node and new node record whether a node has been visited before. As you enter or leave a branch, mark the wall or the floor somehow. Whenever you come to a new node, choose any branch. If you come to a dead end, return to the previous node. If you travel along a path previously untraveled and reach an old node (it will have marks on at least two of the branches), go back to the node you just left. If you are on a previously traveled path, choose a new branch. If that is not possible, choose a branch that has been traversed only once. This procedure is laborious and will probably take you along a lengthy route, but it avoids all traps.

Suppose you enter a maze, pass a number of nodes without taking notes or marking your path and then realize you are lost. How can you best return to the entrance without moving hopelessly deeper into the maze? In 1959 Oystein Ore of Yale University published a procedure for dealing with such a situation.

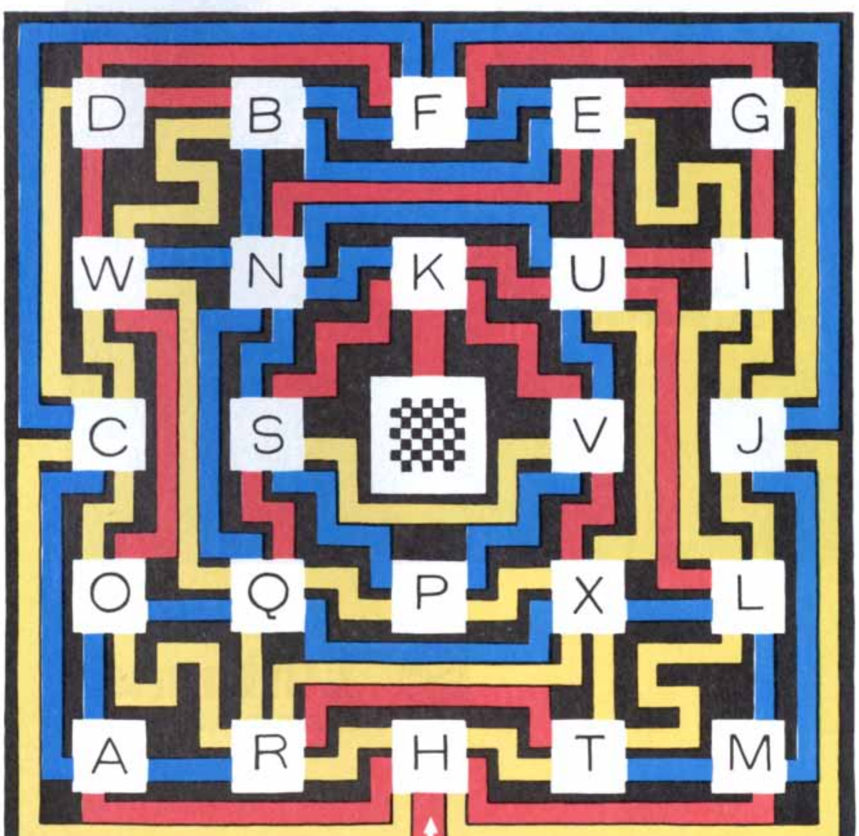
Imagine that in a maze network you are at a point x and not only have lost your way but also have forgotten the number of nodes through which you have passed [*see bottom illustration at left*]. From x explore each branch until you come to another node. As you enter the branch, mark it with a 1. If you come to a node that has new branches, mark the branch you are in with another 1 and return to x. If you come to a dead end, mark the branch as being closed when you return to x. If a branch loops around in such a way as to return to x, mark each end of the branch as being closed.

Next explore each unclosed branch to a distance of one additional node. As you leave x, add another mark of 1 to the entrance of the branch. When you leave the branch at the next node (which you reached on the preceding venture), mark another 1 at this nodal exit point. (The node now has two marks.) As you enter a new branch at the node, mark 1 at its entrance. If the branch is a dead end, return to the node you just left and mark the branch as being closed. If you come to a node that you have already visited from some other branch leading from x, mark each end of the branch you are then in as being closed. An example is the branch connecting nodes a and b in the illustration. Return to the node you just left.

When you have completed the explorations from x to a distance of two nodes in all possible directions, return



*"A*maze*ment" by Minotaur Designs*



"Alphabet Soup"

to *x* and begin exploring to a distance of three nodes. Remember to add a 3 to each branch as you enter it and as you leave it. Note that at any distant node you can always determine the direction back to *x* by comparing the marks on the branches: the branch leading back to *x* has the higher number. The illustration represents explorations to a distance of three nodes. You will regain the entrance when you extend the explorations to a distance of four nodes.

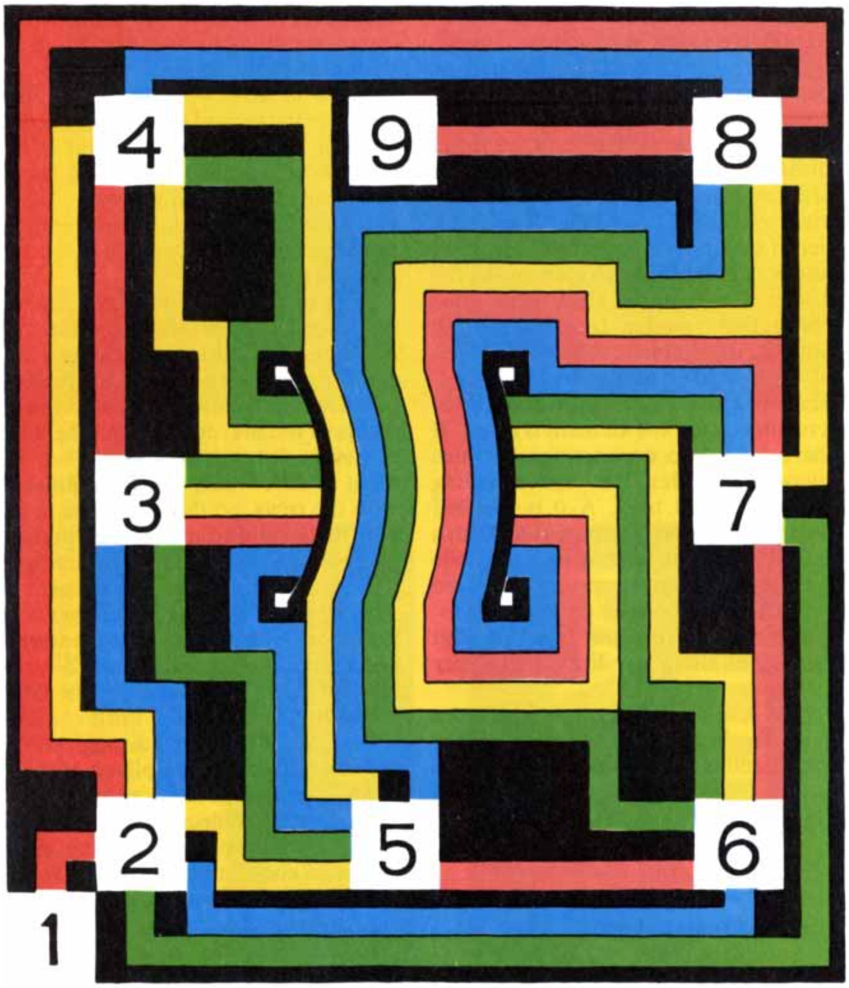
Minotaur Designs is a small company that builds full-size mazes in England and other countries. Adrian Fisher, one of its partners, sent me drawings of three of the company's unique color mazes. The first maze, called "A*maze*ment," was originally laid out in Epsom for an exhibition. You can follow the drawing of it at the left with a finger or a pencil. You enter the maze on the red path leading to node *R* and attempt to reach the goal at the center by way of a minimum of nodes. At each node you must change colors. For example, if you reach node *I* by the blue path, you cannot leave it on another blue path.

The maze design contains many more nodes than the eight indicated by letters. For example, node *I* is actually two distinct nodes, one if you enter it on a red branch and another if you enter it on a blue branch. If you draw a network for this maze, you must include one node for the blue *I* and another node for the red *I*.

The second color maze is called "Alphabet Soup." It too requires a change of colors at each node. Beginning on the red path to node *H*, what is the lowest number of nodes needed to reach the goal in the center? This maze is cleverly designed. If you attempt to work backward from the goal, a standard technique among maze enthusiasts, you quickly lose your way. The interior of the maze, marked by the square defined by *O, D, G* and *L*, is difficult to breach. Many paths extending from the entrance loop back to the entrance. After I had figured out how to reach the interior, I discovered several direct routes with 11 nodes (counting the goal and considering *H* as the first node), but a direct route of 10 nodes took longer to discover. I think it is the minimal route.

The third maze, called "The Giant's Bridge," imposes a different restriction at the nodes. You are required to choose a new branch according to the sequence red, blue, yellow, green. If you enter a node on a red branch, you must leave on a blue one. Green calls for red and so on.

The center of the maze contains a



"The Giant's Bridge"

bridge and an underpass of branches. What is the lowest number of nodes needed to travel from node 1 to the goal of node 9? Fisher's solution (in the form of a network) is shown on the next page. Note that like the other color mazes, the network contains many more nodes than are readily apparent in the maze itself.

The properties of networks were initially outlined in 1735 by the eminent Swiss mathematician Leonhard Euler. Consider any connected network. A node is said to be odd or even according to the number of branches joining it. A route is any consecutive series of branches in which no branch is traversed twice. A reentrant route ends where it begins. A unicursal route goes through the entire network without repeating a branch.

Euler set forth four general rules concerning networks. (1) The number of odd nodes must be even or zero. (2) If a network has no odd nodes, it can be traveled unicursally by beginning at any node. Moreover, any such route is reentrant. (3) If a network has only two odd nodes, it can be traveled unicursally by a route that begins at one of them and ends at the other one. Any route that begins at an even node, however, cannot traverse the network unicursally. (4) Any network that has more than two odd nodes cannot be traveled unicursally by a route. It can be fully explored by several routes without traveling over a branch more than once. If it has $2n$ odd nodes, it can be fully explored in n routes.

Examples of the rules appear in the upper illustration on page 147. In the first network the number of odd nodes is even. If you begin at one of the two odd nodes, you can travel unicursally through the entire network to end at the other odd node. If you begin at the one even node, you need two routes to explore the full network. The second network has an extra branch. Again the number of odd nodes is even. Since there are now more than two odd nodes, a unicursal route is impossible. Exploration of the full network requires at least two routes. Two examples are shown.

Often, but not always, a maze has an odd node at the entrance and another one at the goal. If all the other nodes are even, you can travel from the entrance, through the entire maze and to the goal without going through any branch more than once. If there is even a single additional odd node, however, you will have to travel over at least one branch twice in order to explore the entire maze.

The study of networks, now called graph theory, is widespread in mathematics, electrical engineering, computer processing, route designing and many other fields. Graph theory involves networks of the types applicable to mazes with one exception: it does not allow a branch to leave a node and then loop back to that node. A maze loop of that kind can nonetheless be altered to fit graph theory by inserting an artificial node into the loop. In a maze this node would be a trivial one, involving no decision other than one to quit the current direction in a branch and return to the previous node before the next node is discovered.

Graph theory offers an elegant way to tackle a maze in which you are required to find a minimal route. You begin by constructing a matrix of the connections between successive nodes. The maze network shown in the lower illustration on the opposite page has eight nodes; its matrix is a square with eight elements on a side. The number of connections between successive nodes is entered as an element in the matrix. For example, since there is one connection running from node 1 to node 2, the element with coordinates of 1 (vertical) and 2 (horizontal) is 1. Since you could also travel from node 2 to node 1, the 2–1 element is also 1. If there were two connections between successive nodes, the corresponding element would be 2. A 0 is inserted at all the empty elements. Call this matrix M. Note that it is symmetric around a diagonal running from the top left to the bottom right. The symmetry results from the fact that you can travel along any branch in either direction.

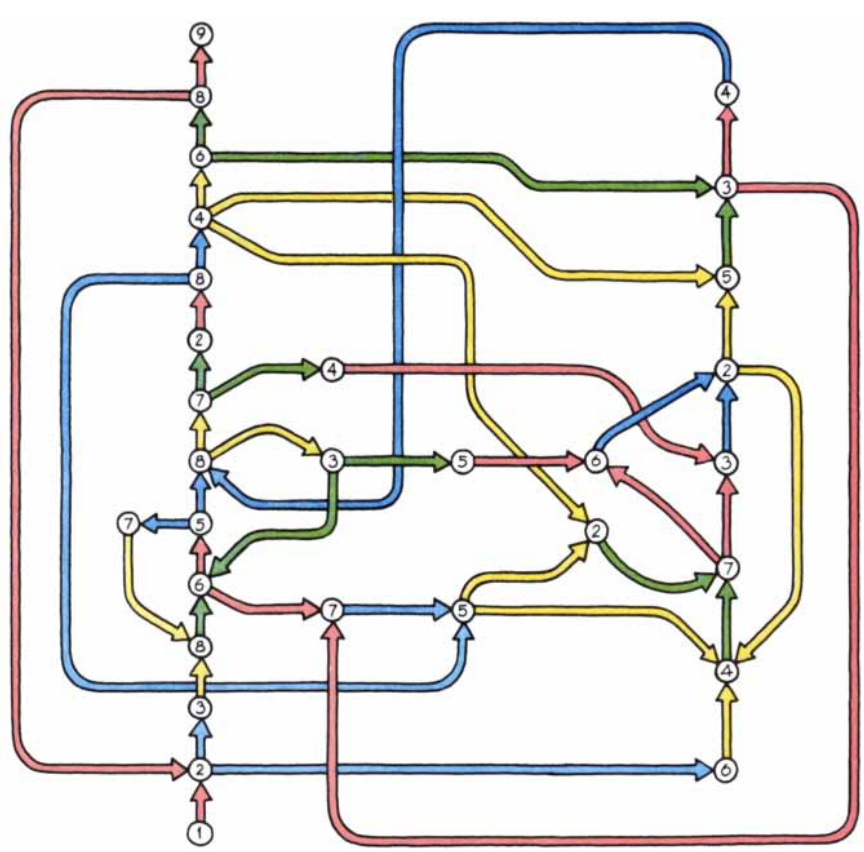
If M is multiplied by itself to yield a new matrix M^2, you can determine which nodes are connected by a route consisting of two branches. The calculation proceeds as follows. Multiply the first element in the first column by the first element in the first row. Then multiply the second element in the first column by the second element in the first row. Continue multiplying corresponding elements in this way. When you have finished multiplying, add the products and enter the result as the 1–1 element of M^2.

Next deal with the first column and the second row. Multiply corresponding elements, add the products and enter the result as the 1–2 element of M^2. Then multiply corresponding elements of the first column and the third row and enter the result as the 1–3 element of M^2. When you have finished with the rows, go through them again with the second column. When the second column and the first row are multiplied, the result is the 2–1 element of M^2. When corresponding elements in the second column and the second row are multiplied, the result is the 2–2 element of M^2. Continue down the rows and then proceed to the third column. When you have worked through all the columns, you have completed M^2.

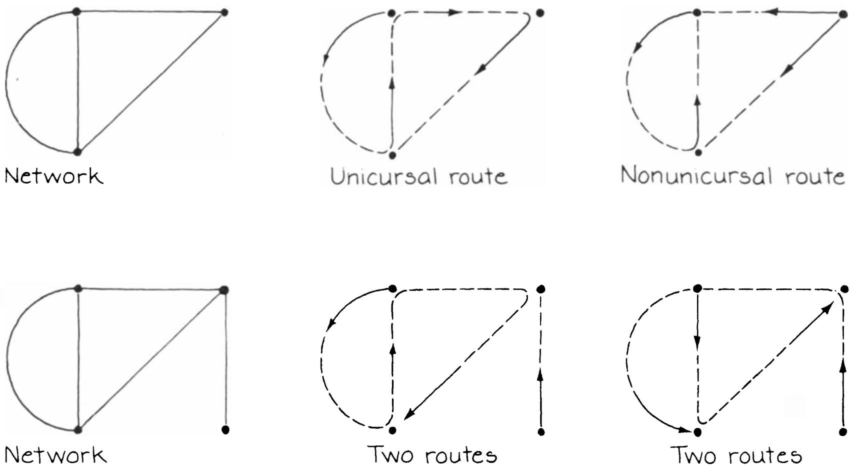
The elements of M^2 lying on the line of symmetry reflect the number of ways you can move from any given node to a connecting node and back to the initial node, thus traveling over a branch twice. The rest of the elements involve travel from one node to another by a route of two branches. Element 2–4 is 1, indicating that there is only one route connecting nodes 2 and 4 by two branches. Element 2–5 is 2, indicating that there are two routes connecting nodes 2 and 5 by two branches each. Is there a route that leads from the entrance node 1 to the goal node 5 by means of two branches? No, there is not, because the 1–5 element in the matrix is 0.

By multiplying M^2 by M to yield M^3, you can generate a matrix that counts the number of ways you can go from one node to another by a route of three branches. For example, the 1–4 element is 1, indicating that there is only one route of three branches linking node 1 with node 4: 1–2–3–4. Some routes are convoluted. For example, of the six routes linking nodes 2 and 3 by three branches, one is 2–3–4–3. Is there a three-branch route linking nodes 1 and 5? Yes, there is: this time the 1–5 element is not 0. The value of 2 indicates that there are two routes consisting of three branches by which it is possible to reach the goal from the entrance node.

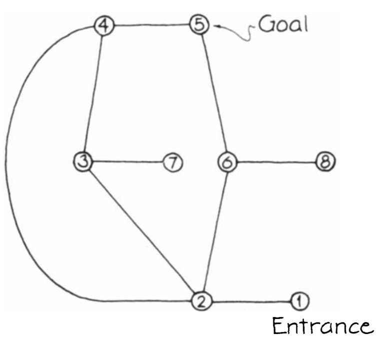
Matrix analysis can be applied to more complex mazes that are not readily solvable by sight. All the nodes



A network for "The Giant's Bridge"



Routes that fully explore networks



M	1	2	3	4	5	6	7	8
1	0	1	0	0	0	0	0	0
2	1	0	1	1	0	1	0	0
3	0	1	0	1	0	0	1	0
4	0	1	1	0	1	0	0	0
5	0	0	0	1	0	1	0	0
6	0	1	0	0	1	0	0	1
7	0	0	1	0	0	0	0	0
8	0	0	0	0	0	1	0	0

M^2	1	2	3	4	5	6	7	8
1	1	0	1	1	0	1	0	0
2	0	4	1	1	2	0	1	1
3	1	1	3	1	1	1	0	0
4	1	1	1	3	0	2	1	0
5	0	2	1	0	2	0	0	1
6	1	0	1	2	0	3	0	0
7	0	1	0	1	0	0	1	0
8	0	1	0	0	1	0	0	1

M^3	1	2	3	4	5	6	7	8
1	0	4	1	1	2	0	1	1
2	4	2	6	7	1	7	1	0
3	1	6	2	5	2	2	3	1
4	1	7	5	2	5	1	1	2
5	2	1	2	5	0	5	1	0
6	0	7	2	1	5	0	1	3
7	1	1	3	1	1	1	0	0
8	1	0	1	2	0	3	0	0

A matrix representing a maze network

are numbered and the number of connections between successive nodes is listed in a matrix. The power of the matrix is raised until a number appears in the element corresponding to a link between the entrance node and the goal node. The power of the matrix is the number of branches in the minimal route for the maze. The matrix does not reveal where direct routes are, but it can indicate whether a direct route you may already have found is the minimal route.

If a maze has one-way branches, its matrix is modified. For example, if you can move from node 4 to node 3 but not in the other direction, the element 4–3 is 1 but the element 3–4 is 0. If you enjoy the matrix approach to a maze, you might like to see whether the minimal route in "Alphabet Soup" consists of 10 nodes as I asserted and whether there is only one such route. A copy of a paperbound book, *A Celebration of Mazes,* is available from Minotaur Designs for £3 in the U.K. and Europe (42 Brampton Road, St. Albans, Hertfordshire AL1 4PT, U.K.) or for $9 in North America (247 Montgomery Steet, Jersey City, N.J. 07302). Readers in other regions should inquire at the U.K. address.

ANNUAL INDEX

AUTHORS

Abbott, Allan V., Alec N. Brooks and David Gordon Wilson. HUMAN-POWERED WATERCRAFT; December, page 120.
Ahlquist, Jon E., and Charles G. Sibley. RECONSTRUCTING BIRD PHYLOGENY BY COMPARING DNA'S; February, page 82.
Anderson, Dana Z. OPTICAL GYROSCOPES; April, page 94.
Anstis, Stuart M., and Vilayanur S. Ramachandran. THE PERCEPTION OF APPARENT MOTION; June, page 102.
Aubert, James H., Andrew M. Kraynik and Peter B. Rand. AQUEOUS FOAMS; May, page 74.
Baer, Eric. ADVANCED POLYMERS; October, page 178.
Baer, William C. THE SHADOW MARKET IN HOUSING; November, page 29.
Bergner, Lawrence, Mickey S. Eisenberg, Alfred P. Hallstrom and Richard O. Cummins. SUDDEN CARDIAC DEATH; May, page 37.
Betz, A. Lorris, and Gary W. Goldstein. THE BLOOD-BRAIN BARRIER; September, page 74.
Blaustein, Andrew R., and Richard K. O'Hara. KIN RECOGNITION IN TADPOLES; January, page 108.
Borgia, Gerald. SEXUAL SELECTION IN BOWERBIRDS; June, page 92.
Bowen, Kent H. ADVANCED CERAMICS; October, page 168.
Bowsher, John, Don Smithers and Klaus Wogram. PLAYING THE BAROQUE TRUMPET; April, page 108.
Bowyer, Stuart, and Francesco Paresce. THE SUN AND THE INTERSTELLAR MEDIUM; September, page 92.
Brandt, John C., and Malcolm B. Niedner, Jr. THE STRUCTURE OF COMET TAILS; January, page 48.
Brooks, Alec N., Allan V. Abbott and David Gordon Wilson. HUMAN-POWERED WATERCRAFT; December, page 120.
Brou, Philippe, Thomas R. Sciascia, Lynette Linden and Jerome Y. Lettvin. THE COLORS OF THINGS; September, page 84.
Burns, Jack O. VERY LARGE STRUCTURES IN THE UNIVERSE; July, page 38.
Bush, Mitchell, Stephen J. O'Brien and David E. Wildt. THE CHEETAH IN GENETIC PERIL; May, page 84.
Cantin, Marc, and Jacques Genest. THE HEART AS AN ENDOCRINE GLAND; February, page 76.
Carter, William E., and Douglas S.

Robertson. STUDYING THE EARTH BY VERY-LONG-BASELINE INTERFEROMETRY; November, page 46.
Cech, Thomas R. RNA AS AN ENZYME; November, page 64.
Chaudhari, Praveen. ELECTRONIC AND MAGNETIC MATERIALS; October, page 136.
Chou, Tsu-Wei, Roy L. McCullough and R. Byron Pipes. COMPOSITES; October, page 192.
Claassen, Richard S., and Louis A. Girifalco. MATERIALS FOR ENERGY UTILIZATION; October, page 102.
Clark, Joel P., and Merton C. Flemings. ADVANCED MATERIALS AND THE ECONOMY; October, page 50.
Compton, W. Dale, and Norman A. Gjostein. MATERIALS FOR GROUND TRANSPORTATION; October, page 92.
Coutant, Charles C. THERMAL NICHES OF STRIPED BASS; August, page 98.
Craxton, R. Stephen, Robert L. McCrory and John M. Soures. PROGRESS IN LASER FUSION; August, page 68.
Crutchfield, James P., J. Doyne Farmer, Norman H. Packard and Robert S. Shaw. CHAOS; December, page 46.
Cummins, Richard O., Mickey S. Eisenberg, Lawrence Bergner and Alfred P. Hallstrom. SUDDEN CARDIAC DEATH; May, page 37.
DeVries, Arthur L., and Joseph T. Eastman. ANTARCTIC FISHES; November, page 106.
Divone, Louis V., and Peter M. Moretti. MODERN WINDMILLS; June, page 110.
Dreesman, Gordon R., Ronald C. Kennedy and Joseph L. Melnick. ANTI-IDIOTYPES AND IMMUNITY; July, page 48.
Eastman, Joseph T., and Arthur L. DeVries. ANTARCTIC FISHES; November, page 106.
Eisenberg, Mickey S., Lawrence Bergner, Alfred P. Hallstrom and Richard O. Cummins. SUDDEN CARDIAC DEATH; May, page 37.
Evans, Michael L., Randy Moore and Karl-Heinz Hasenstein. HOW ROOTS RESPOND TO GRAVITY; December, page 112.
Factor, Stephen M., Thomas F. Robinson and Edmund H. Sonnenblick. THE HEART AS A SUCTION PUMP; June, page 84.
Farmer, J. Doyne, James P. Crutchfield, Norman H. Packard and Robert S. Shaw. CHAOS; December, page 46.
Fine, Alan. TRANSPLANTATION IN THE CENTRAL NERVOUS SYSTEM; August, page 52.
Finke, Ronald A. MENTAL IMAGERY AND THE VISUAL SYSTEM; March, page 88.
Flemings, Merton C., and Joel P. Clark. ADVANCED MATERIALS AND THE ECONOMY; October, page 50.
Foley, Vernard, and Werner Soedel. LEONARDO'S CONTRIBUTIONS TO THEORETICAL MECHANICS; September, page 108.
Foster, Kenneth R., and Arthur W. Guy. THE MICROWAVE PROBLEM; September, page 32.
Fuller, Robert A., and Jonathan J. Rosen. MATERIALS FOR MEDICINE; October, page 118.
Gallo, Robert C. THE FIRST HUMAN RETROVIRUS; December, page 88.
Geldard, Frank A., and Carl E. Sherrick. SPACE, TIME AND TOUCH; July, page 90.
Genest, Jacques, and Marc Cantin. THE HEART AS AN ENDOCRINE GLAND; February, page 76.
Gingerich, Owen. ISLAMIC ASTRONOMY; April, page 74.
Girifalco, Louis A., and Richard S. Claassen. MATERIALS FOR ENERGY UTILIZATION; October, page 102.
Gjostein, Norman A., and W. Dale Compton. MATERIALS FOR GROUND TRANSPORTATION; October, page 92.
Goldanskii, Vitalii I. QUANTUM CHEMICAL REACTIONS IN THE DEEP COLD; February, page 46.
Goldstein, Gary W., and A. Lorris Betz. THE BLOOD-BRAIN BARRIER; September, page 74.
Gosselin, Edward A., and Lawrence S. Lerner. GALILEO AND THE SPECTER OF BRUNO; November, page 126.
Gowlett, John A. J., and Robert E. M. Hedges. RADIOCARBON DATING BY ACCELERATOR MASS SPECTROMETRY; January, page 100.
Green, Michael B. SUPERSTRINGS; September, page 48.
Guy, Arthur W., and Kenneth R. Foster. THE MICROWAVE PROBLEM; September, page 32.
Gwinner, Eberhard. INTERNAL RHYTHMS IN BIRD MIGRATION; April, page 84.
Haber, Howard E., and Gordon L. Kane. IS NATURE SUPERSYMMETRIC?; June, page 52.
Haber, Ralph Norman. FLIGHT SIMULATION; July, page 96.
Haberle, Robert M. THE CLIMATE OF MARS; May, page 54.
Hadley, Neil F. THE ARTHROPOD CUTICLE; July, page 104.
Hakomori, Sen-itiroh. GLYCOSPHINGOLIPIDS; May, page 44.
Hallstrom, Alfred P., Mickey S. Eisenberg, Lawrence Bergner and Richard O. Cummins. SUDDEN CARDIAC DEATH; May, page 37.
Halperin, Bertrand I. THE QUANTIZED HALL EFFECT; April, page 52.
Halstead, Paul, Glynis Jones, Kenneth Wardle and Diana Wardle. CROP STORAGE AT ASSIROS; March, page 96.
Hamilton, J. H., and J. A. Maruhn. EXOTIC ATOMIC NUCLEI; July, page 80.
Hammond, Norman. THE EMERGENCE OF MAYA CIVILIZATION; August, page 106.
Hasenstein, Karl-Heinz, Michael L. Evans and Randy Moore. HOW ROOTS RESPOND TO GRAVITY; December, page 112.
Hedges, Robert E. M., and John A. J. Gowlett. RADIOCARBON DATING BY ACCELERATOR MASS SPECTROMETRY; January, page 100.
Herbert, Sandra. DARWIN AS A GEOLOGIST; May, page 116.
Hinman, C. Wiley. POTENTIAL NEW CROPS; July, page 32.
Hones, Jr., Edward W. THE EARTH'S MAGNETOTAIL; March, page 40.
Hoskin, Michael. WILLIAM HERSCHEL AND THE MAKING OF MODERN ASTRONOMY; February, page 106.
Hynes, Richard O. FIBRONECTINS; June, page 42.
Jackson, J. David, Maury Tigner and Stanley Wojcicki. THE SUPERCONDUCTING SUPERCOLLIDER; March, page 66.
Jones, Glynis, Kenneth Wardle, Paul Halstead and Diana Wardle. CROP STORAGE AT ASSIROS; March, page 96.
Kane, Gordon L., and Howard E. Haber. IS NATURE SUPERSYMMETRIC?; June, page 52.
Kappler, John, and Philippa Marrack. THE *T* CELL AND ITS RECEPTOR; February, page 36.
Karplus, Martin, and J. Andrew McCammon. THE DYNAMICS OF PROTEINS; April, page 42.
Kear, Bernard H. ADVANCED METALS; October, page 158.
Kennedy, Ronald C., Joseph L. Melnick and Gordon R. Dreesman. ANTI-IDIOTYPES AND IMMUNITY; July, page 48.
Krauss, Lawrence M. DARK MATTER IN THE UNIVERSE; December, page 58.
Kraynik, Andrew M., James H. Aubert and Peter B. Rand. AQUEOUS FOAMS; May, page 74.
Kyle, Chester R. ATHLETIC CLOTHING; March, page 104.
Laeser, Richard P., William I. McLaughlin and Donna M. Wolff. ENGINEERING VOYAGER 2'S ENCOUNTER WITH URANUS; November, page 36.
Lagercrantz, Hugo, and Theodore A. Slotkin. THE "STRESS" OF BEING BORN; April, page 100.
Larson, Eric D., Marc H. Ross and Robert H. Williams. BEYOND THE ERA OF MATERIALS; June, page 34.
Lawn, Richard M., and Gordon A. Vehar. THE MOLECULAR GENETICS OF HEMOPHILIA; March, page 48.
Lerner, Lawrence S., and Edward A. Gosselin. GALILEO AND THE SPECTER OF BRUNO; November, page 126.
Lester, Richard K. RETHINKING NUCLEAR POWER; March, page 31.
Lettvin, Jerome Y., Philippe Brou, Thomas R. Sciascia and Lynette Linden. THE COLORS OF THINGS; Septem-

ber, page 84.
Liedl, Gerald L. THE SCIENCE OF MATERIALS; October, page 126.
Light, Donald W. CORPORATE MEDICINE FOR PROFIT; December, page 38.
Linden, Lynette, Philippe Brou, Thomas R. Sciascia and Jerome Y. Lettvin. THE COLORS OF THINGS; September, page 84.
Mangelsdorf, Paul C. THE ORIGIN OF CORN; August, page 80.
Marrack, Philippa, and John Kappler. THE *T* CELL AND ITS RECEPTOR; February, page 36.
Maruhn, J. A., and J. H. Hamilton. EXOTIC ATOMIC NUCLEI; July, page 80.
Masland, Richard H. THE FUNCTIONAL ARCHITECTURE OF THE RETINA; December, page 102.
Mayo, John S. MATERIALS FOR INFORMATION AND COMMUNICATION; October, page 58.
McCammon, J. Andrew, and Martin Karplus. THE DYNAMICS OF PROTEINS; April, page 42.
McCrory, Robert L., R. Stephen Craxton and John M. Soures. PROGRESS IN LASER FUSION; August, page 68.
McCullough, Roy L., Tsu-Wei Chou and R. Byron Pipes. COMPOSITES; October, page 192.
McLaughlin, William I., Richard P. Laeser and Donna M. Wolff. ENGINEERING VOYAGER 2'S ENCOUNTER WITH URANUS; November, page 36.
Melnick, Joseph L., Ronald C. Kennedy and Gordon R. Dreesman. ANTI-IDIOTYPES AND IMMUNITY; July, page 48.
Molnar, Peter. THE STRUCTURE OF MOUNTAIN RANGES; July, page 70.
Moore, Randy, Michael L. Evans and Karl-Heinz Hasenstein. HOW ROOTS RESPOND TO GRAVITY; December, page 112.
Moretti, Peter M., and Louis V. Divone. MODERN WINDMILLS; June, page 110.
Mutter, John C. SEISMIC IMAGES OF PLATE BOUNDARIES; February, page 66.
Nelson, David R. QUASICRYSTALS; August, page 42.
Niedner, Jr., Malcolm B., and John C. Brandt. THE STRUCTURE OF COMET TAILS; January, page 48.
Niklas, Karl J. COMPUTER-SIMULATED PLANT EVOLUTION; March, page 78.
Nincic, Miroslav. CAN THE U.S. TRUST THE U.S.S.R.?; April, page 33.
O'Brien, Stephen J., David E. Wildt and Mitchell Bush. THE CHEETAH IN GENETIC PERIL; May, page 84.
O'Hara, Richard K., and Andrew R. Blaustein. KIN RECOGNITION IN TADPOLES; January, page 108.
Packard, Norman H., James P. Crutchfield, J. Doyne Farmer and Robert S. Shaw. CHAOS; December, page 46.
Paresce, Francesco, and Stuart Bowyer. THE SUN AND THE INTERSTELLAR MEDIUM; September, page 92.
Pepper, David M. APPLICATIONS OF OPTICAL PHASE CONJUGATION; January, page 74.
Phillips, Thomas G., and David B. Rutledge. SUPERCONDUCTING TUNNEL DETECTORS IN RADIO ASTRONOMY; May, page 96.
Pipes, R. Byron, Tsu-Wei Chou and Roy L. McCullough. COMPOSITES; October, page 192.
Prados, John, Joel S. Wit and Michael J. Zagurek, Jr. THE STRATEGIC NUCLEAR FORCES OF BRITAIN AND FRANCE; August, page 33.
Protzen, Jean-Pierre. INCA STONEMASONRY; February, page 94.
Ramachandran, Vilayanur S., and Stuart M. Anstis. THE PERCEPTION OF APPARENT MOTION; June, page 102.
Ramage, Colin S. EL NIÑO; June, page 76.
Rand, Peter B., James H. Aubert and Andrew M. Kraynik. AQUEOUS FOAMS; May, page 74.
Richardson, Joyce R. BRACHIOPODS; September, page 100.
Robertson, Douglas S., and William E. Carter. STUDYING THE EARTH BY VERY-LONG-BASELINE INTERFEROMETRY; November, page 46.
Robinson, Thomas F., Stephen M. Factor and Edmund H. Sonnenblick. THE HEART AS A SUCTION PUMP; June, page 84.
Rona, Peter A. MINERAL DEPOSITS FROM SEA-FLOOR HOT SPRINGS; January, page 84.
Rosen, Jonathan J., and Robert A. Fuller. MATERIALS FOR MEDICINE; October, page 118.
Rosenthal, Gerald A. THE CHEMICAL DEFENSES OF HIGHER PLANTS; January, page 94.
Ross, Marc H., Eric D. Larson and Robert H. Williams. BEYOND THE ERA OF MATERIALS; June, page 34.
Rouvray, Dennis H. PREDICTING CHEMISTRY FROM TOPOLOGY; September, page 40.
Rowell, J. M. PHOTONIC MATERIALS; October, page 146.
Rutledge, David B., and Thomas G. Phillips. SUPERCONDUCTING TUNNEL DETECTORS IN RADIO ASTRONOMY; May, page 96.
Sachs, Leo. GROWTH, DIFFERENTIATION AND THE REVERSAL OF MALIGNANCY; January, page 40.
Sciascia, Thomas R., Philippe Brou, Lynette Linden and Jerome Y. Lettvin. THE COLORS OF THINGS; September, page 84.
Shaw, Robert S., James P. Crutchfield, J. Doyne Farmer and Norman H. Packard. CHAOS; December, page 46.
Sherrick, Carl E., and Frank A. Geldard. SPACE, TIME AND TOUCH; July, page 90.
Sibley, Charles G., and Jon E. Ahlquist. RECONSTRUCTI‸G BIRD PHYLOGENY BY COMPARING DNA'S; February, page 82.
Slotkin, Theodore A., and Hugo Lagercrantz. THE "STRESS" OF BEING BORN; April, page 100.
Smithers, Don, Klaus Wogram and John Bowsher. PLAYING THE BAROQUE TRUMPET; April, page 108.
Soedel, Werner, and Vernard Foley. LEONARDO'S CONTRIBUTIONS TO THEORETICAL MECHANICS; September, page 108.
Sonnenblick, Edmund H., Thomas F. Robinson and Stephen M. Factor. THE HEART AS A SUCTION PUMP; June, page 84.
Soures, John M., R. Stephen Craxton and Robert L. McCrory. PROGRESS IN LASER FUSION; August, page 68.
Steinberg, Morris A. MATERIALS FOR AEROSPACE; October, page 66.
Tigner, Maury, J. David Jackson and Stanley Wojcicki. THE SUPERCONDUCTING SUPERCOLLIDER; March, page 66.
Treisman, Anne. FEATURES AND OBJECTS IN VISUAL PROCESSING; November, page 114B.
Tschirley, Fred H. DIOXIN; February, page 29.
Van Allen, James A. SPACE SCIENCE, SPACE TECHNOLOGY AND THE SPACE STATION; January, page 32.
Vehar, Gordon A., and Richard M. Lawn. THE MOLECULAR GENETICS OF HEMOPHILIA; March, page 48.
Veltman, Martinus J. G. THE HIGGS BOSON; November, page 76.
Wardle, Diana, Glynis Jones, Kenneth Wardle and Paul Halstead. CROP STORAGE AT ASSIROS; March, page 96.
Wardle, Kenneth, Glynis Jones, Paul Halstead and Diana Wardle. CROP STORAGE AT ASSIROS; March, page 96.
Watts, Carol Martin, and Donald J. Watts. A ROMAN APARTMENT COMPLEX; December, page 132.
Watts, Donald J., and Carol Martin Watts. A ROMAN APARTMENT COMPLEX; December, page 132.
Wildt, David E., Stephen J. O'Brien and Mitchell Bush. THE CHEETAH IN GENETIC PERIL; May, page 84.
Williams, George E. THE SOLAR CYCLE IN PRECAMBRIAN TIME; August, page 88.
Williams, Robert H., Eric D. Larson and Marc H. Ross. BEYOND THE ERA OF MATERIALS; June, page 34.
Wilson, David Gordon, Alec N. Brooks and Allan V. Abbott. HUMAN-POWERED WATERCRAFT; December, page 120.
Wit, Joel S., John Prados and Michael J. Zagurek, Jr. THE STRATEGIC NUCLEAR FORCES OF BRITAIN AND FRANCE; August, page 33.
Wogram, Klaus, Don Smithers and John Bowsher. PLAYING THE BAROQUE TRUMPET; April, page 108.

Wojcicki, Stanley, J. David Jackson and Maury Tigner. THE SUPERCONDUCTING SUPERCOLLIDER; March, page 66.
Wolff, Donna M., Richard P. Laeser and William I. McLaughlin. ENGINEERING VOYAGER 2'S ENCOUNTER WITH URANUS; November, page 36.
Zagurek, Jr., Michael J., John Prados and Joel S. Wit. THE STRATEGIC NUCLEAR FORCES OF BRITAIN AND FRANCE; August, page 33.
Zvelebil, Marek. POSTGLACIAL FORAGING IN THE FORESTS OF EUROPE; May, page 104.

ARTICLES

ANTI-IDIOTYPES AND IMMUNITY, by Ronald C. Kennedy, Joseph L. Melnick and Gordon R. Dreesman; July, page 48.
ARTHROPOD CUTICLE, THE, by Neil F. Hadley; July, page 104.
ASSIROS, CROP STORAGE AT, by Glynis Jones, Kenneth Wardle, Paul Halstead and Diana Wardie; March, page 96.
ATHLETIC CLOTHING, by Chester R. Kyle; March, page 104.
ATOMIC NUCLEI, EXOTIC, by J. H. Hamilton and J. A. Maruhn; July, page 80.
BIRD MIGRATION, INTERNAL RHYTHMS IN, by Eberhard Gwinner; April, page 84.
BIRD PHYLOGENY BY COMPARING DNA'S, RECONSTRUCTING, by Charles G. Sibley and Jon E. Ahlquist; February, page 82.
BLOOD-BRAIN BARRIER, THE, by Gary W. Goldstein and A. Lorris Betz; September, page 74.
BORN, THE "STRESS" OF BEING, by Hugo Lagercrantz and Theodore A. Slotkin; April, page 100.
BOWERBIRDS, SEXUAL SELECTION IN, by Gerald Borgia; June, page 92.
BRACHIOPODS, by Joyce R. Richardson; September, page 100.
CARDIAC DEATH, SUDDEN, by Mickey S. Eisenberg, Lawrence Bergner, Alfred P. Hallstrom and Richard O. Cummins; May, page 37.
CERAMICS, ADVANCED, by H. Kent Bowen; October, page 168.
CHAOS, by James P. Crutchfield, J. Doyne Farmer, Norman H. Packard and Robert S. Shaw; December, page 46.
CHEETAH IN GENETIC PERIL, THE, by Stephen J. O'Brien, David E. Wildt and Mitchell Bush; May, page 84.
CHEMICAL REACTIONS IN THE DEEP COLD, QUANTUM, by Vitalii I. Goldanskii; February, page 46.
CHEMISTRY FROM TOPOLOGY, PREDICTING, by Dennis H. Rouvray; September, page 40.
COLORS OF THINGS, THE, by Philippe Brou, Thomas R. Sciascia, Lynette Linden and Jerome Y. Lettvin; September, page 84.
COMET TAILS, THE STRUCTURE OF, by John C. Brandt and Malcolm B. Niedner, Jr.; January, page 48.
COMPOSITES, by Tsu-Wei Chou, Roy L. McCullough and R. Byron Pipes; October, page 192.
CORN, THE ORIGIN OF, by Paul C. Mangelsdorf; August, page 80.
CORPORATE MEDICINE FOR PROFIT, by Donald W. Light; December, page 38.
CROPS, POTENTIAL NEW, by C. Wiley Hinman; July, page 32.
DARK MATTER IN THE UNIVERSE, by Lawrence M. Krauss; December, page 58.
DARWIN AS A GEOLOGIST, by Sandra Herbert; May, page 116.
DATING BY ACCELERATOR MASS SPECTROMETRY, RADIOCARBON, by Robert E. M. Hedges and John A. J. Gowlett; January, page 100.
DIOXIN, by Fred H. Tschirley; February, page 29.
EL NIÑO, by Colin S. Ramage; June, page 76.
FIBRONECTINS, by Richard O. Hynes; June, page 42.
FISHES, ANTARCTIC, by Joseph T. Eastman and Arthur L. DeVries; November, page 106.
FLIGHT SIMULATION, by Ralph Norman Haber; July, page 96.
FOAMS, AQUEOUS, by James H. Aubert, Andrew M. Kraynik and Peter B. Rand; May, page 74.
FORAGING IN THE FORESTS OF EUROPE, POSTGLACIAL, by Marek Zvelebil; May, page 104.
FUSION, PROGRESS IN LASER, by R. Stephen Craxton, Robert L. McCrory and John M. Soures; August, page 68.
GALILEO AND THE SPECTER OF BRUNO, by Lawrence S. Lerner and Edward A. Gosselin; November, page 126.
GLYCOSPHINGOLIPIDS, by Sen-itiroh Hakomori; May, page 44.
GYROSCOPES, OPTICAL, by Dana Z. Anderson; April, page 94.
HEART AS AN ENDOCRINE GLAND, THE, by Marc Cantin and Jacques Genest; February, page 76.
HEART AS A SUCTION PUMP, THE, by Thomas F. Robinson, Stephen M. Factor and Edmund H. Sonnenblick; June, page 84.
HEMOPHILIA, THE MOLECULAR GENETICS OF, by Richard M. Lawn and Gordon A. Vehar; March, page 48.
HERSCHEL AND THE MAKING OF MODERN ASTRONOMY, WILLIAM, by Michael Hoskin; February, page 106.
HIGGS BOSON, THE, by Martinus J. G. Veltman; November, page 76.
HOUSING, THE SHADOW MARKET IN, by William C. Baer; November, page 29.
IMAGERY AND THE VISUAL SYSTEM, MENTAL, by Ronald A. Finke; March, page 88.
INCA STONEMASONRY, by Jean-Pierre Protzen; February, page 94.
INTERFEROMETRY, STUDYING THE EARTH BY VERY-LONG-BASELINE, by William E. Carter and Douglas S. Robertson; November, page 46.
INTERSTELLAR MEDIUM, THE SUN AND THE, by Francesco Paresce and Stuart Bowyer; September, page 92.
ISLAMIC ASTRONOMY, by Owen Gingerich; April, page 74.
KIN RECOGNITION IN TADPOLES, by Andrew R. Blaustein and Richard K. O'Hara; January, page 108.
LARGE STRUCTURES IN THE UNIVERSE, VERY, by Jack O. Burns; July, page 38.
LEONARDO'S CONTRIBUTIONS TO THEORETICAL MECHANICS, by Vernard Foley and Werner Soedel; September, page 108.
MAGNETOTAIL, THE EARTH'S, by Edward W. Hones, Jr.; March, page 40.
MALIGNANCY, GROWTH, DIFFERENTIATION AND THE REVERSAL OF, by Leo Sachs; January, page 40.
MARS, THE CLIMATE OF, by Robert M. Haberle; May, page 54.
MATERIALS AND THE ECONOMY, ADVANCED, by Joel P. Clark and Merton C. Flemings; October, page 50.
MATERIALS, BEYOND THE ERA OF, by Eric D. Larson, Marc H. Ross and Robert H. Williams; June, page 34.
MATERIALS, ELECTRONIC AND MAGNETIC, by Praveen Chaudhari; October, page 136.
MATERIALS FOR AEROSPACE, by Morris A. Steinberg; October, page 66.
MATERIALS FOR ENERGY UTILIZATION, by Richard S. Claassen and Louis A. Girifalco; October, page 102.
MATERIALS FOR GROUND TRANSPORTATION, by W. Dale Compton and Norman A. Gjostein; October, page 92.
MATERIALS FOR INFORMATION AND COMMUNICATION, by John S. Mayo; October, page 58.
MATERIALS FOR MEDICINE, by Robert A. Fuller and Jonathan J. Rosen; October, page 118.
MATERIALS, PHOTONIC, by J. M. Rowell; October, page 146.
MATERIALS, THE SCIENCE OF, by Gerald L. Liedl; October, page 126.
MAYA CIVILIZATION, THE EMERGENCE OF, by Norman Hammond; August, page 106.
METALS, ADVANCED, by Bernard H. Kear; October, page 158.
MICROWAVE PROBLEM, THE, by Kenneth R. Foster and Arthur W. Guy; September, page 32.
MOUNTAIN RANGES, THE STRUCTURE OF, by Peter Molnar; July, page 70.
NERVOUS SYSTEM, TRANSPLANTATION IN THE CENTRAL, by Alan Fine; August, page 52.
NUCLEAR FORCES OF BRITAIN AND FRANCE, THE STRATEGIC, by John Prados, Joel S. Wit and Michael J. Za-

gurek, Jr.; August, page 33.
NUCLEAR POWER, RETHINKING, by Richard K. Lester; March, page 31.
PERCEPTION OF APPARENT MOTION, THE, by Vilayanur S. Ramachandran and Stuart M. Anstis; June, page 102.
PHASE CONJUGATION, APPLICATIONS OF OPTICAL, by David M. Pepper; January, page 74.
PLANT EVOLUTION, COMPUTER-SIMULATED, by Karl J. Niklas; March, page 78.
PLANTS, THE CHEMICAL DEFENSES OF HIGHER, by Gerald A. Rosenthal; January, page 94.
POLYMERS, ADVANCED, by Eric Baer; October, page 178.
PROTEINS, THE DYNAMICS OF, by Martin Karplus and J. Andrew McCammon; April, page 42.
QUANTIZED HALL EFFECT, THE, by Bertrand I. Halperin; April, page 52.
QUASICRYSTALS, by David R. Nelson; August, page 42.
RETINA, THE FUNCTIONAL ARCHITECTURE OF THE, by Richard H. Masland; December, page 102.
RETROVIRUS, THE FIRST HUMAN, by Robert C. Gallo; December, page 88.
RNA AS AN ENZYME, by Thomas R. Cech; November, page 64.
ROMAN APARTMENT COMPLEX, A, by Donald J. Watts and Carol Martin Watts; December, page 132.
ROOTS RESPOND TO GRAVITY, HOW, by Michael L. Evans, Randy Moore and Karl-Heinz Hasenstein; December, page 112.
SEA-FLOOR HOT SPRINGS, MINERAL DEPOSITS FROM, by Peter A. Rona; January, page 84.
SEISMIC IMAGES OF PLATE BOUNDARIES, by John C. Mutter; February, page 66.
SOLAR CYCLE IN PRECAMBRIAN TIME, THE, by George E. Williams; August, page 88.
SPACE SCIENCE, SPACE TECHNOLOGY AND THE SPACE STATION, by James A. Van Allen; January, page 32.
STRIPED BASS, THERMAL NICHES OF, by Charles C. Coutant; August, page 98.
SUPERCOLLIDER, THE SUPERCONDUCTING, by J. David Jackson, Maury Tigner and Stanley Wojcicki; March, page 66.
SUPERSTRINGS, by Michael B. Green; September, page 48.
SUPERSYMMETRIC?, IS NATURE, by Howard E. Haber and Gordon L. Kane; June, page 52.
T CELL AND ITS RECEPTOR, THE, by Philippa Marrack and John Kappler; February, page 36.
TOUCH, SPACE, TIME AND, by Frank A. Geldard and Carl E. Sherrick; July, page 90.
TRUMPET, PLAYING THE BAROQUE, by Don Smithers, Klaus Wogram and John Bowsher; April, page 108.
TUNNEL DETECTORS IN RADIO ASTRONOMY, SUPERCONDUCTING, by Thomas G. Phillips and David B. Rutledge; May, page 96.
U.S.S.R.?, CAN THE U.S. TRUST THE, by Miroslav Nincic; April, page 33.
VISUAL PROCESSING, FEATURES AND OBJECTS IN, by Anne Treisman; November, page 114B.
VOYAGER 2'S ENCOUNTER WITH URANUS, ENGINEERING, by Richard P. Laeser, William I. McLaughlin and Donna M. Wolff; November, page 36.
WATERCRAFT, HUMAN-POWERED, by Alec N. Brooks, Allan V. Abbott and David Gordon Wilson; December, page 120.
WINDMILLS, MODERN, by Peter M. Moretti and Louis V. Divone; June, page 110.

COMPUTER RECREATIONS

Caricaturist and a whimsical tour of face space, The compleat computer; October, page 20.
Chess program) is dead, long live the king (a chess machine), The king (a; February, page 13.
Computer graphics at Pixar, Of fractal mountains, graftal plants and other; December, page 14.
Flight of fancy over a deserted data base, A sublime; July, page 16.
Hypercubes induces four-dimensional dementia, A program for rotating; April, page 14.
I.Q. tests, How a pair of dull-witted programs can look like geniuses on; March, page 14.
Magic and illusion by computer, Digital prestidigitation: the fine art of; August, page 17.
Phylogenies of the Paleozoic and the fortunes of English family names, Branching; May, page 16.
Star clusters are achieved with a computer telescope, How close encounters with; January, page 16.
Star Trek emerges from the underground to a place in the home-computer arcade; November, page 14.
Wallpaper for the mind: computer images that are almost, but not quite, repetitive; September, page 14.

MATHEMATICAL GAMES

Net on a checkerboard and other puzzles of the forest, Casting a; June, page 16.

THE AMATEUR SCIENTIST

Aluminum metal was initially purified back in 1886, Retracing the steps by which; August, page 116.
Color in a black-and-white grating, Methods and optics of perceiving; March, page 112.
Cracks in a surface look intricately random but actually develop rather systematically; October, page 204.
Holograms, unlike conventional ones, can be observed in ordinary light, Rainbow; September, page 114.
Maze so bewildering that the explorer must rely on a map, Mirrors make a; June, page 120.
Maze without becoming lost or confused, Methods for going through a; December, page 140.
Particle scattering; experiments in zero gravity, A homemade device for testing; February, page 114.
Polarimeter can analyze optically active compounds, An inexpensive homemade; January, page 120.
Retroreflector, a mirror that removes distortion from a light beam, Wonders with the; April, page 118.
Shape is put to work running an engine, Wire that "remembers" its; May, page 124.
Three-dimensional vision, The hyperscope and the pseudoscope aid experiments on; November, page 134.
Water when it is freezing or melting, Exotic patterns appear in; July, page 114.

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION (required by 39 U.S.C. 3685). *1A.* Title of publication: Scientific American. *1B.* Publication number: 00368733. *2.* Date of filing: September 29, 1986. *3.* Frequency of issue: monthly. *3A.* Number of issues published annually: 12. *3B.* Annual subscription price: U.S. and its possessions, 1 year $24; all other countries, 1 year $33. *4.* Complete mailing address of known office of publication: 415 Madison Avenue, New York, N.Y. 10017. *5.* Complete mailing address of the headquarters or general business offices of the publisher: 415 Madison Avenue, New York, N.Y. 10017. *6.* Full names and complete mailing address of publisher, editor, and managing editor: Publisher, Harry Myers, 415 Madison Avenue, New York, N.Y. 10017. Editor, Jonathan Piel, 415 Madison Avenue, New York, N.Y. 10017. Managing Editor, none. *7.* Owner: Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Holtzbrinck Publishing Holdings Limited Partnership, 521 Fifth Avenue, New York, N.Y. 10175. *8.* Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities: none. *9.* For completion by nonprofit organizations authorized to mail at special rates (Section 423.12, DMM only). The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: Not applicable to this organization. *10.* Extent and nature of circulation: *A.* Total number of copies (net press run): average number of copies each issue during preceding 12 months, 760,405; actual number of copies of single issue published nearest to filing date, 714,320. *B.* Paid and/or requested circulation: *1.* Sales through dealers and carriers, street vendors and counter sales: average number of copies each issue during preceding 12 months, 118,409; actual number of copies of single issue published nearest tc filing date, 123, 011. *2.* Mail subscription (paid and/or requested): average number of copies each issue during preceding 12 months, 479,368; actual number of copies of single issue published nearest to filing date, 445,861. *C.* Total paid and/or requested circulation (sum of *10B1* and *110B2*): average number of copies each issue during preceding 12 months, 597,777; actual number of copies of single issue published nearest to filing date, 568,872. *D.* Free distribution by mail, carrier or other means, samples, complimentary, and other free copies: average number of copies each issue during preceding 12 months, 4,810; actual number of copies of single issue published nearest to filing date, 3,766. *E.* Total distribution (sum of *C* and *D*): average number of copies each issue during preceding 12 months, 602,587; actual number of copies of single issue published nearest to filing date, 572,638. *F.* Copies not distributed: *1.* Office use, leftover, unaccounted, spoiled after printing: average number of copies each issue during preceding 12 months, 12, 516; actual number of copies of single issue published nearest to filing date, 10,964. *2.* Return from news agents: average number of copies each issue during preceding 12 months, 145,302; actual number of copies of single issue published nearest to filing date, 130,718. *G.* Total (sum of *E, F1* and *2*—should equal net press run shown in *A*): average number of copies each issue during preceding 12 months, 760,405; actual number of copies of single issue published nearest to filing date, 714,320. *11.* I certify that the statements made by me above are correct and complete. (Signed) Jonathan Piel, Editor.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

DISTRIBUTED RAY TRACING. Robert L. Cook, Thomas Porter and Loren Carpenter in *Computer Graphics,* Vol. 18, No. 3, pages 137–145; July, 1984.

3-D FRACTALS. Michiel van de Panne in *Creative Computing,* Vol. 11, No. 7, pages 78–82; July, 1985.

CORPORATE MEDICINE FOR PROFIT

FOR-PROFIT ENTERPRISE IN HEALTH CARE. Edited by Bradford H. Gray. Institute of Medicine, National Academy Press, 1986.

THE COMPARATIVE ECONOMIC PERFORMANCE OF INVESTOR-OWNED CHAIN AND NOT-FOR-PROFIT HOSPITALS. J. Michael Watt, Robert A. Derzon, Steven C. Renn, Carl J. Schramm, James S. Hahn and George D. Pillari in *The New England Journal of Medicine,* Vol. 314, No. 2, pages 89–96; January 9, 1986.

A NEW LOOK AT NONPROFITS: HEALTH CARE POLICY IN A COMPETITIVE AGE. Theodore R. Marmor, Mark Schlesinger and Richard W. Smithey in *Yale Journal on Regulation,* Vol. 3, No. 2, pages 313–349; Spring, 1986.

CHAOS

GEOMETRY FROM A TIME SERIES. N. H. Packard, J. P. Crutchfield, J. D. Farmer and R. S. Shaw in *Physical Review Letters,* Vol. 45, No. 9, pages 712–716; September 1, 1980.

DYNAMICS: THE GEOMETRY OF BEHAVIOR. Ralph Abraham and Chris Shaw. Aerial Press, P.O. Box 1360, Santa Cruz, Calif. 95061; 1982–85.

DETERMINISTIC CHAOS: AN INTRODUCTION. H. G. Schuster. VCH Publishers, Inc., 1984.

THE EUDAEMONIC PIE: OR WHY WOULD ANYONE PLAY ROULETTE WITHOUT A COMPUTER IN HIS SHOE? Thomas A. Bass. Houghton-Mifflin, 1985.

DIMENSIONS AND ENTROPIES IN CHAOTIC SYSTEMS. Edited by G. Mayer-Kress. Springer-Verlag, 1986.

DARK MATTER IN THE UNIVERSE

THE LARGE-SCALE STRUCTURE OF THE UNIVERSE. P. J. E. Peebles. Princeton University Press, 1980.

DARK MATTER AND INFLATION. Lawrence M. Krauss in *General Relativity and Gravitation,* Vol. 17, No. 1, pages 89–94; January, 1985.

INNER SPACE, OUTER SPACE. Edited by E. W. Kolb, M. S. Turner, D. Lindley, K. Olive and D. Seckel. University of Chicago Press, 1986.

SOLAR SYSTEM CONSTRAINTS AND SIGNATURES FOR DARK-MATTER CANDIDATES. Lawrence M. Krauss, Mark Srednicki and Frank Wilczek in *Physical Review D,* Vol. 33, No. 8, pages 2079–2083; April 15, 1986.

THE MEAN MASS DENSITY OF THE UNIVERSE. P. J. E. Peebles in *Nature,* Vol. 321, No. 6065, pages 27–32; May 1, 1986.

THE FIRST HUMAN RETROVIRUS

THE FAMILY OF HUMAN LYMPHOTROPIC RETROVIRUSES CALLED HTLV: HTLV-I IN ADULT T-CELL LEUKEMIA (ATL), HTLV-II IN HAIRY CELL LEUKEMIAS, AND HTLV-III IN AIDS. R. C. Gallo in *Retroviruses in Human Lymphoma/Leukemia,* edited by M. Miwa et al. Japan Scientific Societies Press, 1985.

HUMAN T-CELL LEUKEMIA/LYMPHOMA VIRUSES ATL AND AIDS. R. C. Gallo and W. A. Blattner in *Important Advances in Oncology, 1985,* edited by Vincent T. DeVita, Samuel Hellman and Steven A. Rosenberg. J. B. Lippincott Co., 1985.

HUMAN T-CELL LEUKEMIA VIRUSES (HTLV): A UNIQUE FAMILY OF PATHOGENIC RETROVIRUSES. Samuel Broder and Robert C. Gallo in *Annual Review of Immunology,* Vol. 3, pages 321–336; 1985.

HUMAN T-LYMPHOTROPIC RETROVIRUSES. Flossie Wong-Staal and Robert C. Gallo in *Nature,* Vol. 317, No. 6036, pages 395–403; October 3, 1985.

THE FUNCTIONAL ARCHITECTURE OF THE RETINA

THE STRUCTURE OF THE RETINA. Santiago Ramón y Cajal. Charles C. Thomas, 1972.

DENDRITIC TERRITORIES OF CAT RETINAL GANGLION CELLS. H. Wässle, L. Peichl and B. B. Boycott in *Nature,* Vol. 292, No. 5821, pages 344–345; July 23, 1981.

THE SHAPE AND ARRANGEMENT OF THE CHOLINERGIC NEURONS IN THE RABBIT RETINA. M. Tauchi and R. H. Masland in *Proceedings of the Royal Society of London, Series B,* Vol. 223, No. 1230, pages 101–119; November 22, 1984.

THE RETINA—FROM MOLECULES TO NETWORKS. *Trends in Neuroscience,* Vol. 9, No. 5; May, 1986.

HOW ROOTS RESPOND TO GRAVITY

REVERSIBLE LOSS OF GRAVITROPIC SENSITIVITY IN MAIZE ROOTS AFTER TIP APPLICATION OF CALCIUM CHELATORS. June S. Lee, Timothy J. Mulkey and Michael L. Evans in *Science,* Vol. 220, No. 4604, pages 1375–1376; June 24, 1983.

REGULATION OF ROOT DEVELOPMENT. L. J. Feldman in *Annual Review of Plant Physiology,* Vol. 35, pages 223–242; 1984.

HOW ROOTS PERCEIVE AND RESPOND TO GRAVITY. Randy Moore and Michael L. Evans in *American Journal of Botany,* Vol. 73, No. 4, pages 574–587; April, 1986.

HUMAN-POWERED WATERCRAFT

BICYCLING SCIENCE. Frank R. Whitt and David G. Wilson. The MIT Press, 1982.

THE FLYING FISH HYDROFOIL. Alec N. Brooks in *Human Power,* Vol. 3, No. 2, pages 1, 7–8; Winter, 1984.

PROCEEDINGS OF THE THIRD INTERNATIONAL HUMAN POWERED VEHICLE SCIENTIFIC SYMPOSIUM. Edited by Allan V. Abbott. International Human Powered Vehicle Association, P.O. Box 51255, Indianapolis, Ind. 46251; 1986.

A ROMAN APARTMENT COMPLEX

THE SECRETS OF ANCIENT GEOMETRY—AND ITS USE. Tons Brunés. Rhodos International Science and Art Publishers, Copenhagen, 1967.

ROMAN OSTIA. Russell Meiggs. Oxford University Press, 1973.

SACRED GEOMETRY: PHILOSOPHY AND PRACTICE. Robert Lawlor. Crossroad Publishing Co., 1982.

THE AMATEUR SCIENTIST

AN EXCURSION IN LABYRINTHS. Oystein Ore in *The Mathematics Teacher,* Vol. 52, No. 5, pages 367–370; May, 1959.

GRAPHS, MODELS, AND FINITE MATHEMATICS. Joseph Malkevitch and Walter Meyer. Prentice-Hall, Inc., 1974.

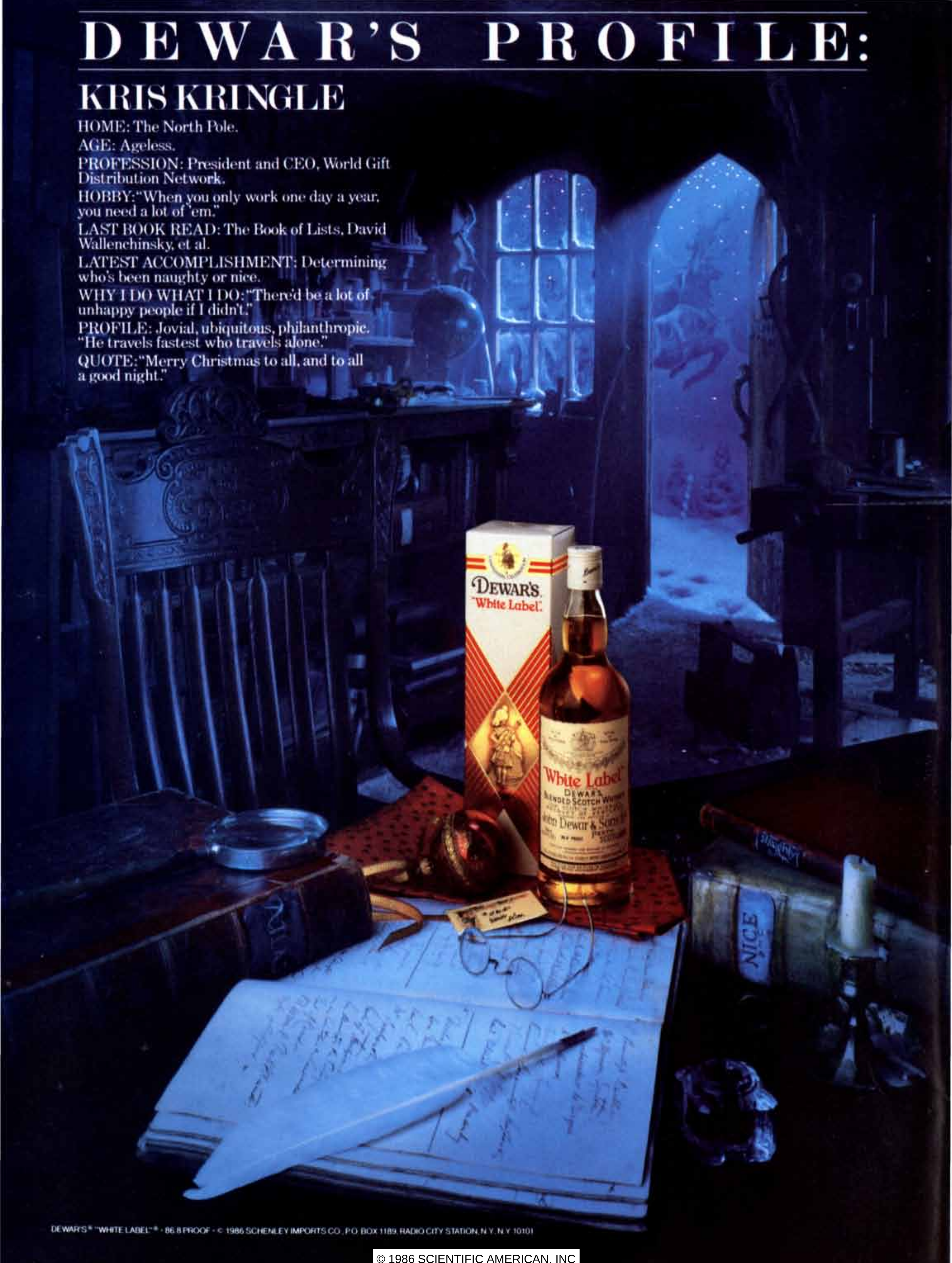
THE LAZARE DIAMOND™

WHY THE IDEAL CUT
IS THE MOST BRILLIANT IDEA
IN DIAMONDS.

Lazare Diamonds are cut to ideal proportions, which is why they capture and release light with more brilliance than any other diamond. For only when a diamond is ideal cut will it display all its light and sparkle. Judge for yourself by asking one of our knowledgeable jewelers to compare the Lazare Diamond with any other.

For the name of our quality jeweler in your area, just call our toll-free number: 1-800-543-8800. Ext 283.

The Lazare Diamond – Setting the Standard for brilliance.™


DEWAR'S PROFILE:
KRIS KRINGLE
HOME: The North Pole.
AGE: Ageless.
PROFESSION: President and CEO, World Gift Distribution Network.
HOBBY: "When you only work one day a year, you need a lot of 'em."
LAST BOOK READ: The Book of Lists, David Wallenchinsky, et al.
LATEST ACCOMPLISHMENT: Determining who's been naughty or nice.
WHY I DO WHAT I DO: "There'd be a lot of unhappy people if I didn't."
PROFILE: Jovial, ubiquitous, philanthropic. "He travels fastest who travels alone."
QUOTE: "Merry Christmas to all, and to all a good night."
DEWAR'S
"White Label"
White Label
DEWAR'S
BLENDED SCOTCH WHISKY
John Dewar & Sons Ltd
NICE
DEWAR'S® "WHITE LABEL"® · 86.8 PROOF · © 1986 SCHENLEY IMPORTS CO., P.O. BOX 1189, RADIO CITY STATION, N.Y., N.Y. 10101