SCIENTIFIC AMERICAN

VOYAGER 2: TRIUMPH IN SPACE

$2.50

November 1986

MERCURY SABLE.

THE BEAUTY OF IT IS IT'S A WAGON.

For more information, call toll-free ***1-800-822-9292***

Buckle up—together we can save lives.

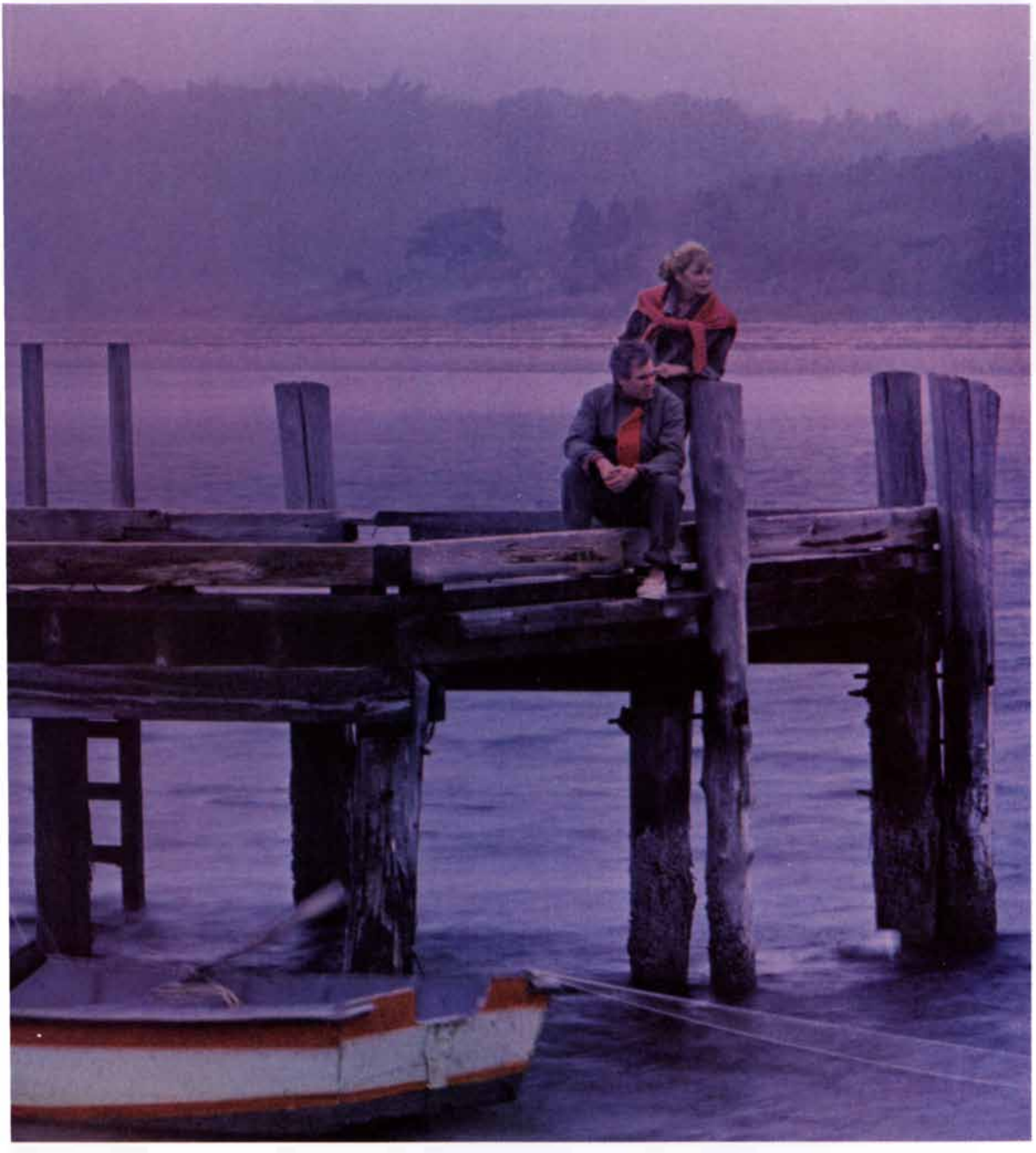
Lincoln-Mercury Division Ford

Time was, if you wanted a great road car and the carrying capacity of a station wagon, you had to buy two cars.

Enter Mercury Sable. It has front-wheel drive, fully independent suspension and a 3.0-liter V-6 engine. All within a shape that makes it the most aerodynamically efficient station wagon built in America today.

Now just because it isn't bulky, doesn't mean it can't carry plenty of bulk. Sable's 60/40 split rear seat lets you fold all or part of the seatback down flat. That gives you a maximum cargo area that's six and a half feet long and more than three feet wide.

Sable. When a wagon is also a great road car, it looks like this.



MERCURY. THE SHAPE YOU WANT TO BE IN.

Established 1845 **SCIENTIFIC AMERICAN** November 1986 Volume 255 Number 5

ARTICLES

DEPARTMENTS

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.

SCIENTIFIC AMERICAN

In Other Languages

LE SCIENZE
L. 3,500/copy L. 35,000/year L. 45,000/(abroad)
Editorial, subscription correspondence:
Le Scienze S.p.A., Via del Lauro, 14,
20121 Milano, Italy
Advertising correspondence:
Publietas, S.p.A., Via Cino de Duca, 5,
20122 Milano, Italy

サイエンス
Y880/copy Y9600/year Y13,000/(abroad)
Editorial, subscription, advertising correspondence:
Nikkei Science, Inc.
No. 9-5, 1-Chome, Otemachi
Chiyoda-ku, Tokyo, Japan

INVESTIGACION Y CIENCIA
450 Ptas/copy 4950Ptas/year $35/(abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica S.A.,
Calabria, 235-239
08029 Barcelona, Spain

POUR LA SCIENCE
25FF/copy 245FF/year 290FF/year (abroad)
Editorial, subscription, advertising correspondence:
Pour la Science S.A.R.L.,
8, rue Férou,
75006 Paris, France

Spektrum DER WISSENSCHAFT
9.80 DM/copy 99 DM/year 112.20 DM/(abroad)
Editorial, subscription correspondence:
Spektrum der Wissenschaft GmbH & Co.
Moenchhofstrasse, 15
D-6900 Heidelberg,
Federal Republic of Germany
Advertising correspondence:
Gesellschaft Für Wirtschaftspublizistik
Kasernenstrasse 67
D-4000 Duesseldorf,
Federal Republic of Germany

科学
1.40RMB/copy 16RMB/year $24/(abroad)
Editorial, subscription correspondence:
ISTIC-Chongqing Branch, P.O. Box 2104,
Chongqing, People's Republic of China

В МИРЕ НАУКИ
2R/copy 24R/year $70/(abroad)
Editorial correspondence:
MIR Publishers
2, Pervy Rizhsky Pereulok
129820 Moscow U.S.S.R.
Subscription correspondence:
Victor Kamkin, Inc.
12224 Parklawn Drive,
Rockville, MD 20852, USA

TUDOMÁNY
98Ft/copy 1,176Ft/year 2,100Ft/(abroad)
Editorial correspondence:
TUDOMÁNY
H-1536 Budapest, Pf 338
Hungary
Subscription correspondence:
"KULTURA"
H-3891 Budapest, Pf. 149
Hungary

مجلة العلوم
1KD/copy 10KD/year $40/(abroad)
Editorial, subscription, advertising correspondence:
MAJALLAT AL-OLOOM
P.O. BOX 20856 Safat,
13069 – Kuwait

Advertising correspondence all editions:
SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue
New York, NY 10017
Telephone: (212) 754-0262 Telex: 236115

THE COVER

The painting on the cover depicts the scene that would have been presented to an observer alongside *Voyager 2* as the spacecraft made its closest approach to Uranus, just after crossing the plane defined by the planet's rings. The event took place on January 24 of this year, and it resulted in a wealth of invaluable data about the planet, its rings and its satellites. To get *Voyager 2* there and to ensure that its scientific instruments would be functioning satisfactorily required a considerable amount of skill and ingenuity (see "Engineering *Voyager 2*'s Encounter with Uranus," by Richard P. Laeser, William I. McLaughlin and Donna M. Wolff, page 36). The spacecraft, which was launched more than nine years ago, is visually dominated by its white 3.7-meter parabolic radio antenna. Extending upward from the base of the antenna is a truss on which various instruments, including cameras, spectrometers and particle detectors, are fastened; below are three electric-power generators and a long, gold-colored boom holding a pair of magnetometers.

THE ILLUSTRATIONS

Cover painting by Ken Hodges

Page	Source
15–16	David Wiseman, University of Western Ontario
17	Johnny Johnson
30	Queen Ann Historical Society (*top*), Cardwell/Thomas & Associates (*bottom*)
31–35	Joan Starwood
37	Jon Brenneis
38–39	Hank Iken
40	Hank Iken (*top*), Jet Propulsion Laboratory/National Aeronautics and Space Administration (*bottom*)
41–43	Hank Iken
44–45	Jet Propulsion Laboratory/National Aeronautics and Space Administration
47	Douglas S. Robertson, National Geodetic Survey
48–54	George Kelvin, Science Graphics
65–74	Thomas C. Moore
77	Hank Iken
78	Edward Bell
79–84	Andrew Christie
107	Joseph T. Eastman, Ohio University, and Arthur L. DeVries, University of Illinois at Urbana-Champaign
108	Tom Prentiss
109	Edward Bell
110–112	Tom Prentiss
114	Joseph T. Eastman, Ohio University, and Arthur L. DeVries, University of Illinois at Urbana-Champaign (*left*); Tom Prentiss (*right*)
115	Jon Brenneis
116–124	Jerome Kuhl
127	Scala/Art Resource
128	Yale University Library
129	Bibliothèque Nationale
130	Scala/Art Resource
131	Quesada/Burke, courtesy of Rare Book Division, Astor, Lenox and Tilden Foundations, New York Public Library
132	Rare Book Division, Astor, Lenox and Tilden Foundations, New York Public Library
134–138	Michael Goodman

You will find this label sewn into clothing worth having in your wardrobe.

Clothing worth having not for the status it will confer upon you, or the instant identity it will impart. But for its spirit. Comfort. Conscientious construction. Value.

You'll find this label sewn into the collar of the Rugby Shirt so tough, it hands down over the years from parent to child. Into the waistband of Cotton Twill Trousers whose silky, sturdy feel recalls service days. On the inside of a Tweed Jacket whose Highland heathers pick you up on days when the world makes you feel a little blue.

In a world of the flimsy and flighty and fly by night, a label like this can be a wonderful thing.

Please send free catalog.
Lands' End Dept. Q-45
Dodgeville, WI 53595

Name _____

Address _____

City _____

State _____ Zip _____

Or call Toll-free:
800-356-4444

LETTERS

To the Editors:

We are pleased to see *Scientific American* publicizing the important phenomenon called El Niño [see "El Niño," by Colin S. Ramage; SCIENTIFIC AMERICAN, June]. El Niño is a term used to designate the occasional large and anomalous warmings in the eastern and central tropical Pacific Ocean, which disrupt the marine ecosystem. These warmings tend to occur from three to seven years apart but with varying intensity.

The sea-surface warmings result from changes in the trade winds over the tropical Pacific and the consequent changes in ocean currents and in the upwelling of cold water from the depths. The anomalously warm sea temperatures, in turn, appear to drive further changes in atmospheric circulation, not only in the Tropics but throughout the global atmosphere. These corresponding variations in the atmosphere are known as the Southern Oscillation. In combination El Niño and the Southern Oscillation represent the best example of a coupling between the atmosphere and the oceans that has a significant effect on short-term climate.

In the 1982–83 El Niño event, for instance, devastating anomalies in weather and short-term climate occurred all over the globe, ranging from drought in Australia, India, Indonesia and Africa to floods in Peru and Ecuador, coastal flooding in California and along the west coast of South America, displaced storm tracks across the U.S. and unusual hurricane tracks in the Pacific. Similar changes, although often less dramatic, had occurred in previous El Niño events.

Meteorologists and oceanographers consider El Niño to be scientifically very challenging, as well as of great economic and practical importance. Moreover, because individual El Niño events evolve over a period of about two years in an established manner, once an El Niño is recognized as being under way there is potential for adding skill to seasonal forecasting of weather and climate.

In recognition of the potential for seasonal forecasting, the economic and social value of increased understanding of El Niño itself and the challenge posed by the scientific questions, a major U.S. and international program known as TOGA, for Tropical Oceans and Global Atmosphere, is under way to investigate interannual variability in those oceans and the atmosphere. TOGA is a decade-long endeavor (1985–94) under the auspices of the World Climate Research Program. Many countries are very interested and actively involved in this program, and at a meeting in Geneva in May about 20 nations committed resources to it. As a result an intensive period of new observations, analyses, computer modeling and research is beginning. Its object is to determine how the observed relations between variations in sea-surface temperatures in the tropical oceans, of which El Niño is the best example, and the associated fluctuations in the global atmospheric circulation can be utilized to give better weather predictions. The optimism expressed for TOGA by scientists and governments throughout the world was not reflected in Ramage's article. Indeed, it failed to mention TOGA at all.

Contrary to the impression Ramage gives, one potential benefit of TOGA for the U.S. is improved skill in forecasting winter temperatures. Since the waves in atmospheric circulation are affected by El Niño, some places get warmer during El Niño events and some get colder. Simple comparisons of winter temperatures with an index of the Southern Oscillation reveal significant negative correlations over the West Coast, Alaska and Canada and significant positive correlations over the southeastern U.S.

In his article Ramage used temperatures in Illinois as representative of relations between El Niño and U.S. winter temperatures. That comparison is quite misleading. Since Illinois is in the transition zone, the lack of a significant correlation there is not surprising and does *not* imply that no useful relations exist!

We hope this letter will help to offset Ramage's negative view of the state of knowledge and instead convey some of the excitement felt by scientists around the world as they enter this challenging but promising avenue of research.

KEVIN E. TRENBERTH

HARRY VAN LOON

National Center
for Atmospheric Research

MARK CANE

Columbia University

J. MICHAEL WALLACE

University of Washington

JOHN YOUNG

University of Wisconsin

OTIS BROWN

University of Miami

GENE RASMUSSON

University of Maryland

PETER WEBSTER

Pennsylvania State University

TIM BARNETT

Scripps Institution of Oceanography

To the Editors:

I am delighted that so many of my distinguished colleagues have called attention to the impressive international programs embodied in TOGA. Since that project has already been widely publicized, however, I chose to discuss individuals and their work rather than plans that were generated by committee.

As designers and justifiers of a big project, my colleagues are officially optimistic. I am therefore not surprised by their statements that "El Niño events evolve over a period of about two years in an established manner" and that the West Coast of the U.S. is one of the places whose winter temperatures are linked to El Niño. The link between the phenomenon and winter temperatures is rather tenuous in California. Seventeen moderate or strong Niños occurred between 1895 and 1983. During the same period California experienced nine cold winters (with temperatures more than one standard deviation below normal) and eight warm winters (with temperatures more than one standard deviation above normal), of which only one (the warm winter of 1939–40) coincided with a Niño.

Far from holding a "negative view of the state of knowledge," I believe even failure of the several confident forecasts of a moderate or strong 1986 Niño could significantly enhance our understanding.

COLIN S. RAMAGE

University of Hawaii at Manoa

EDITOR'S NOTE

Elements of the drawing on page 100 of the July issue were based on a flight simulator developed by the HTM Corporation of Orlando, Fla.

Our car stereos are so remarkable you read about them in the paper every day.

If you want to know what people think of car stereo systems, check the auto classifieds. Very rarely do people list a stereo system as a reason to buy their car. Unless it's a Blaupunkt. The car stereo systems built by Bosch.

50 Years of Firsts.

Blaupunkts have a reputation for being among the finest in the world. And with good reason. They were the first to bring FM stereo and a fully electronic entertainment system into the automobile.

Now Blaupunkt owners in several American cities are being introduced to ARI—our exclusive traffic information system that speeds driving by automatically interrupting a car's radio or cassette player to give important traffic, road and weather conditions.

But through all these innovations, we've never forgotten that people want one fundamental thing from a car stereo: the best sound possible.

So every one of the car stereos we design provides no less than optimum sound performance. No matter how basic or how elaborate the model.

That Same Design Philosophy Applies To Everything We Make.

Bosch anti-lock brakes are engineered to simply stop cars better. Our fuel injection system helps cars go farther, faster and on less fuel than ever before.

Outside the automotive industry, Bosch expertise is at work designing and manufacturing automation systems, packaging machinery, TV studio and video graphics equipment, power tools, home appliances and medical equipment.

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—9000 in R&D alone—who are capable of finding the right answer to the toughest challenges.

Including yours.

ARI® ● BLAUPUNKT

Trademarks of Blaupunkt-Werke GmbH-Bosch Group.

BOSCH

Bringing high technology down to earth.

the hottest idea in turbine engines is eight thousand years old.

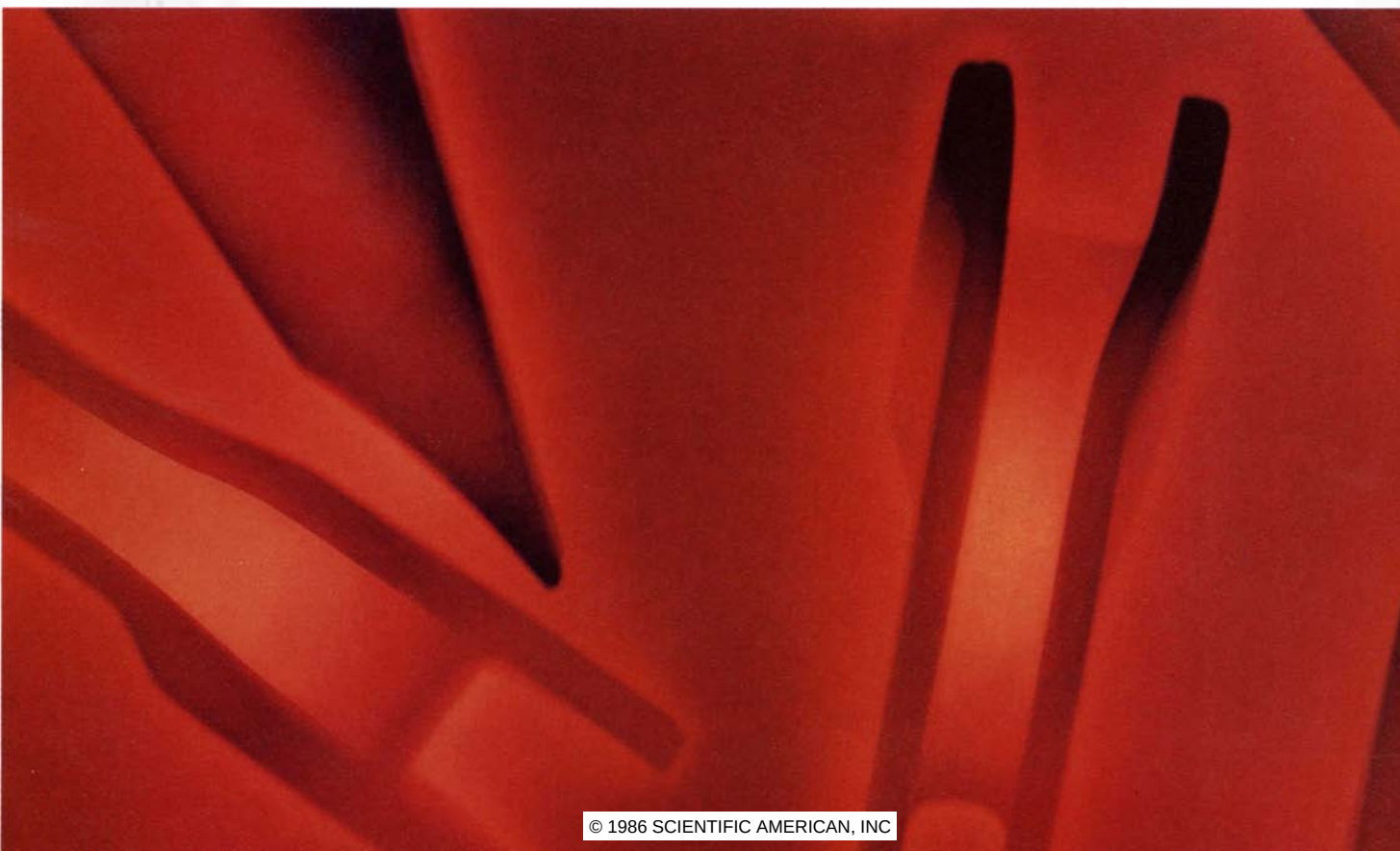
technology: the ancient art of ceramics provides the basis for high temperature components use in Garrett advanced gas turbine engines, from Allied-Signal.

Allied Signal
Technologies

benefit: using ceramics' superior high temperature capability can help engines burn fuel more efficiently, reduce cost of parts, increase life of engine.

future: increasingly, ceramics will replace metals and metal alloys for engine parts, both in aerospace and automotive applications.

advanced ceramics for Garrett turbines. one of the Allied-Signal advanced technologies—part of our growing role in aerospace, electronics, automotive products, and engineered materials.
call 1-800-243-8160.



50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

NOVEMBER, 1936: "R. A. Lyttleton has advanced one of the perplexing problems of cosmogeny by suggesting a way of origin for the planets that avoids the hopeless dynamical difficulties that obsess the older theories. If the sun had originally another star revolving around it—perhaps at about the present distance of Uranus or Neptune—and a third star, passing by in space, collided with the companion, or at least passed exceedingly close to it, it is quite possible that both stars might fly away into space in different directions, leaving masses of ejected matter under the control of the sun's gravitation to condense into the planets."

"With the recent announcement that more than 3,000 people in this country alone are already wearing contact lenses regularly, the invisible, tiny, saucer-shaped lenses, shaped to fit the eye and worn under the eyelids, have now changed from a scientific curiosity to the eye physician's new substitute for spectacles. Although the principle of contact lenses has been known for more than 100 years, during the past two years improvements in the delicate manufacturing processes under which these lenses are ground have been perfected. The chief improvement is a refinement in the method of grinding the inner surface of the little shell so that it presents a continuous gradual curve where it is in contact with the eyeball, which has made for greater comfort."

"Construction of the Albert Canal across Belgium, linking the great port of Antwerp with the nation's industrial center at Liège, is the outstanding task now engaging the engineers and contractors of that enterprising country. When completed next year, as scheduled, the project will have entailed outlays totaling about 80,000,000 dollars. The potential benefits will richly compensate for the money spent."

"With the thought in mind that this is the appropriate time to preserve contemporary records for future generations, Oglethorpe University in Atlanta, in co-operation with SCIENTIFIC AMERICAN, proposes to make available to another civilization now unthought of, and still far in the future, the running story of the life, manners and customs of the present civilization. We propose to collect a complete set of materials that describe and represent our lives and labors, to bury these materials in a secure spot and to preserve them under the guidance and advice of our greatest scientists. Oglethorpe University has selected an ideal spot for this purpose in the basement of a building that now houses its library and executive offices."

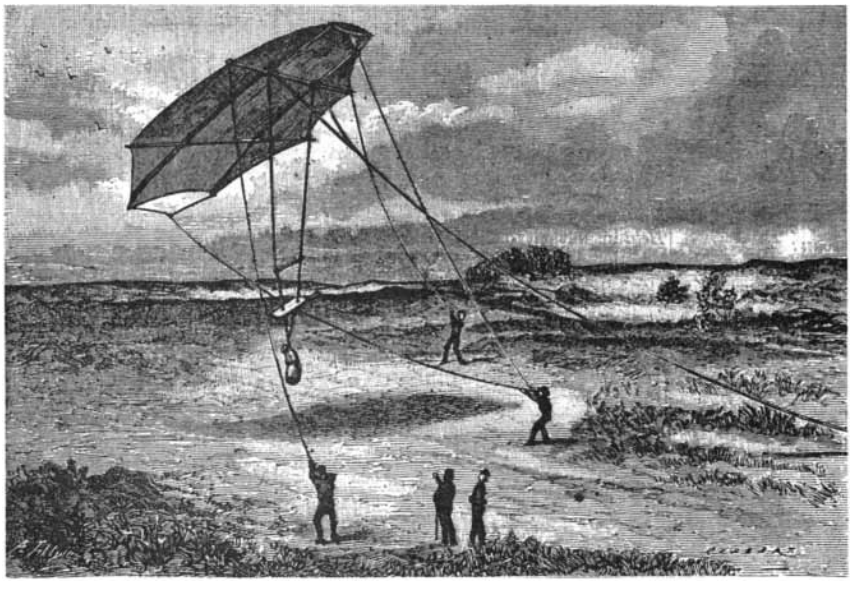
NOVEMBER, 1886: "On October 28 the Statue of Liberty was formally presented to the people of the United States. The public ceremonies in connection therewith constituted one of the greatest pageants of the day. They included a grand parade, a naval demonstration and addresses by President Cleveland and (in French) by Count Senator Ferdinand de Lesseps."

"In a report by Dr. Frantzel of Berlin on immoderate smoking and its effects upon the heart, it is stated that the latter show themselves chiefly by rapid, irregular palpitation of the heart, disturbances in the region of the heart, short breath, languor, sleeplessness, etc. Generally the men smoke full-flavored cigars imported from the Havanas. It is strange that persons consuming cigars of ordinary quality rarely are attacked in that way. The excessive use of cigarettes has not been known to give rise to similar troubles, although it is the cause of complaints of a different nature."

"According to a recent report by the Hon. Geo. C. Tanner, United States Consul at Chemnitz, Germany, the citizens of this country have as yet no adequate idea of the real science of drinking. It has been known and practiced in Europe for centuries, and it is a science, simple as it may appear, when compared with the blind, irrational and suicidal manner of drinking in the United States. This science consists simply in the tardiness of drinking. All drinks are taken sip by sip, a half hour or three-quarters of an hour often being spent in consuming a single glass of beer."

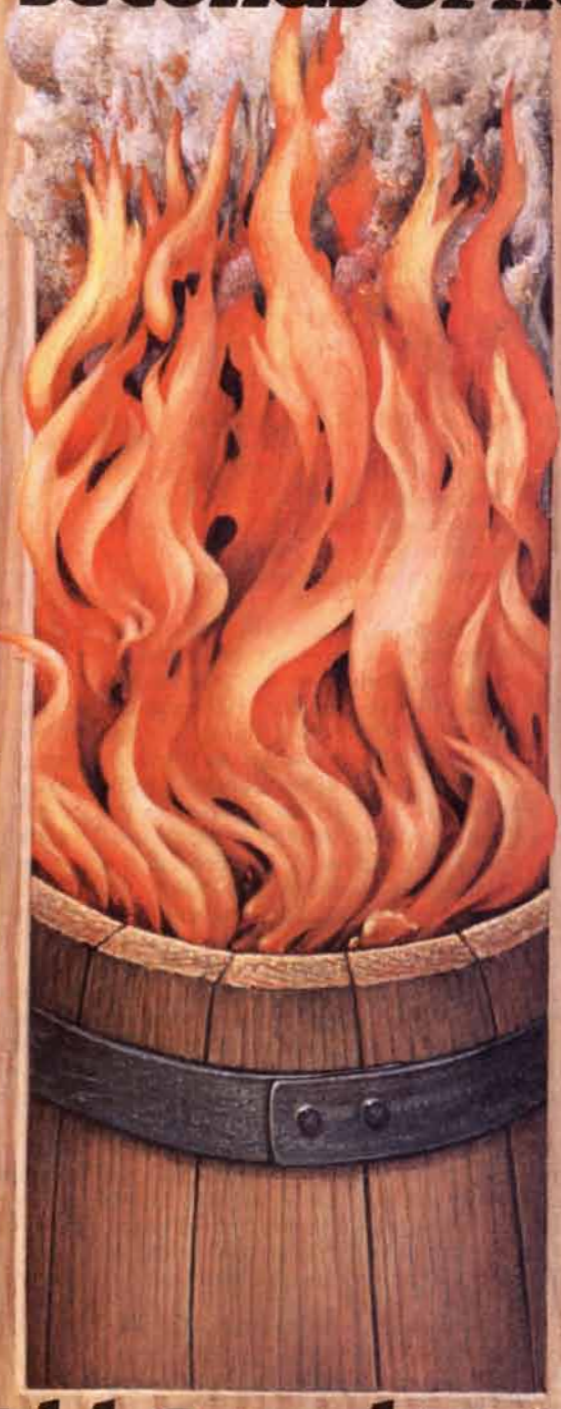
"To produce a small engine that can be operated by the combustion of petroleum is a problem that has received the attention of quite a number of our best inventors. That there is a great demand for small power machines there is no question; and the almost unlimited supply of petroleum, and the low price it is sold for, induce the seeker for a cheap, small power source to turn his thoughts in the direction of petroleum for the fuel."

"After remaining for a long time an object of amusement only, the kite is becoming one of study for the mechanician, who finds in it a means of applying and verifying a number of formulas relating to the resistance of the air. Such experiments in aerodynamics are contributing to the solution of the difficult and complicated problem of flight."



Aerodynamic tests with a large kite

We put every Old Grand-Dad barrel through 45 seconds of hell.



If you've ever watched pure white sugar melt until it becomes delicious brown caramel, you'll appreciate why we're so all-fired particular about the way the white oak barrels we age Old Grand-Dad in are charred.

You see, wood has sugar, too, though it's not the kind you sprinkle on your cornflakes. The intense heat from charring turns some of it into wood-sugar caramel.

As the bourbon slowly ages, it's this reddish caramel that gives Old Grand-Dad its rich, ruddy color and helps create its full-bodied flavor.

Sounds simple.

But it's not.

Each barrel has to be charred to exactly the right depth.

A fraction of an inch either way can actually make the difference between a magnificent bourbon and a mediocre one.

That's why our barrels are charred for 45 seconds—not a tick more or a tock less. And at 1300° F.

We put over 100 years of this kind of skill into the making of every choice, mellow ounce of Old Grand-Dad. No wonder it's the undisputed head of the bourbon family.

Old Grand-Dad

HEAD OF THE BOURBON FAMILY

Kentucky Straight Bourbon Whiskey. 86 Proof. Old Grand-Dad Distillery Co., Frankfort, KY 40601 © 1986 National Distillers

THE AUTHORS

WILLIAM C. BAER ("The Shadow Market in Housing") is associate professor of urban and regional planning at the University of Southern California. He received a master's degree in urban planning from the University of California at Berkeley in 1966 and a doctorate in political science from the University of California at Los Angeles in 1974. Formerly dean of the School of Urban and Regional Planning at U.S.C., Baer currently teaches courses on domestic and international housing.

RICHARD P. LAESER, WILLIAM I. McLAUGHLIN and DONNA M. WOLFF ("Engineering *Voyager 2*'s Encounter with Uranus") have all been deeply involved with the Voyager project in their work at the Jet Propulsion Laboratory. Laeser is project manager, a position he has held since 1981. He previously served as mission director during the two Voyager launches and the Jupiter and Saturn encounters by both spacecraft. He holds two degrees in electrical engineering: a B.S. from the Massachusetts Institute of Technology and an M.S. from the University of Southern California. McLaughlin was manager of flight engineering during the Uranus mission of *Voyager 2*. After getting his Ph.D. in mathematics from the University of California at Berkeley in 1968, he went to work on the Apollo lunar-landing program at Bellcomm, Inc. In 1971 he joined the J.P.L., where he has taken part in numerous space projects including the Viking mission to Mars, the oceanographic satellite *Seasat* and the *Infrared Astronomical Satellite* (*IRAS*). Wolff, who has been at the J.P.L. since 1974, was deputy manager of flight engineering during the Uranus mission. In addition to working on the Voyager project, she has helped to design several space missions, including the *Galileo* Jupiter probe and the *Ulysses* probe to the sun (both of which have now been indefinitely postponed). Wolff got her B.S. in applied mathematics from the California Institute of Technology in 1977 and an M.S. in statistics from Rutgers University in 1979.

WILLIAM E. CARTER and DOUGLAS S. ROBERTSON ("Studying the Earth by Very-Long-Baseline Interferometry") work for the U.S. National Geodetic Survey. Carter began a long career as a geodesist in the U.S. Air Force. After eight years as a geodetic officer, he joined the Air Force Cambridge Research Laboratory in 1969. While he was there he worked on the design, development and operation of a pioneering lunar laser-ranging observatory based in Tucson. In 1973, the year he received his Ph.D. from the University of Arizona, he was appointed director of the Haleakala Observatory of the University of Hawaii, at Manoa. In that capacity he oversaw the design, construction, testing and operation of the observatory. While in Hawaii he also indulged in his favorite hobby: scuba diving and snorkeling. Carter joined the National Geodetic Survey in 1976. Robertson has a B.S. in physics from Principia College and a Ph.D. in earth and planetary science from the Massachusetts Institute of Technology. After getting his doctorate he joined the Computer Sciences Corporation and developed programs for analyzing data from very-long-baseline interferometry. In 1977 he went to the National Geodetic Survey; there he investigates the applications of radio interferometry to problems of the earth's rotation and structure.

THOMAS R. CECH ("RNA as an Enzyme") is professor of chemistry and biochemistry at the University of Colorado at Boulder. He earned a B.A. in 1970 from Grinnell College and did graduate study at the University of California at Berkeley, getting his Ph.D. in 1975. After working in the department of biology at the Massachusetts Institute of Technology as a National Cancer Institute fellow, he joined the faculty at Colorado in 1978.

MARTINUS J. G. VELTMAN ("The Higgs Boson") is John D. MacArthur professor of physics at the University of Michigan. He received his Ph.D. from the University of Utrecht in 1963. That same year he came to the U.S. and joined the Stanford Linear Accelerator Center at Stanford University. In 1966 he was appointed professor at Utrecht and thereafter turned his attention to the problem of renormalization of Yang-Mills fields. In 1981 Veltman moved to Michigan.

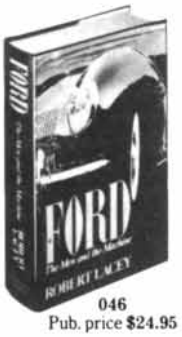
JOSEPH T. EASTMAN and ARTHUR L. DeVRIES ("Antarctic Fishes") are respectively associate professor of anatomy and zoology at Ohio University and professor of physiology at the University of Illinois at Urbana. Eastman did his undergraduate and graduate work at the University of Minnesota, getting his Ph.D. in 1970. He then taught at the University of Oklahoma Medical School for a year before he was appointed assistant professor of anatomical sciences there. He was on the faculty of the Brown University Medical School from 1973 until 1979, when he accepted his current position. Eastman has done field research in Antarctica on three occasions: once to study the anatomy of Antarctic seals and the other times to study the anatomy of Antarctic fishes. DeVries has a bachelor's degree from the University of Montana and a doctorate from Stanford University. It was while he was a graduate student that he first began his research on biological antifreeze compounds in cold-water fishes. From 1971 to 1976 he was associate research physiologist at the Scripps Institution of Oceanography. He then joined the faculty of the University of Illinois at Urbana-Champaign, being appointed full professor in 1984.

ANNE TREISMAN ("Features and Objects in Visual Processing") is professor of psychology at the University of California at Berkeley. Born in Britain, she received her undergraduate education at the University of Cambridge and her graduate training at the University of Oxford. After holding research and fellowship positions in Britain and the U.S., she was appointed professor of psychology at the University of British Columbia in 1978. The research described in her article was supported by the Canadian Institute of Advanced Research and by the Natural Sciences and Engineering Research Council of Canada. Treisman moved to Berkeley this year.

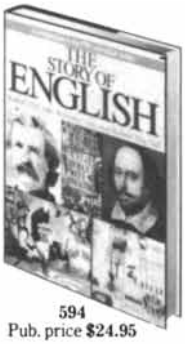
LAWRENCE S. LERNER and EDWARD A. GOSSELIN ("Galileo and the Specter of Bruno") began a collaboration and friendship of long standing when they taught an interdisciplinary course at the California State University at Long Beach, where Lerner is professor of physics and astronomy and Gosselin is professor of history. Lerner, who holds a doctorate from the University of Chicago in solid-state physics, worked for several years in research laboratories in the aerospace and electronics industries before joining the faculty at Long Beach in 1969. He teaches courses in the history of science and in relations between science and society as well as more conventional courses in physics. Gosselin holds a B.A. from Yale University and a Ph.D. from Columbia University. His major research interests are Renaissance and Reformation intellectual and cultural history and the history of science.

Announcing our new "get-acquainted" membership.

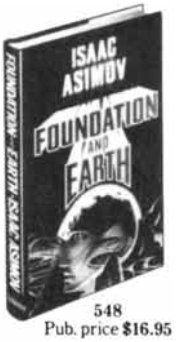
678
Pub. price $25



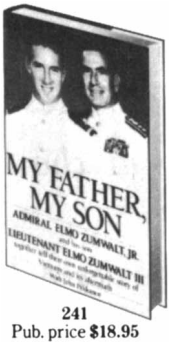
046
Pub. price $24.95



594
Pub. price $24.95

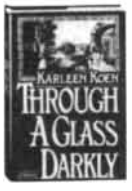


548
Pub. price $16.95

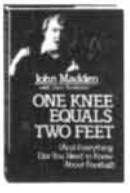


241
Pub. price $18.95

591
Pub. price $22.95

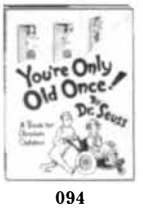


606
Pub. price $19.95

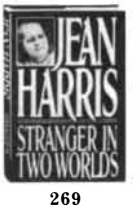


549
Pub. price $16.95

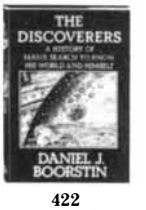
320
Pub. price $27.50



094
Pub. price $10.95

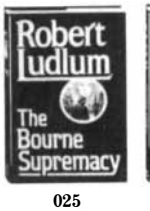


269
Pub. price $18.95

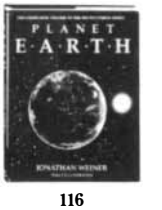


422
Pub. price $25

121
Pub. price $19.95

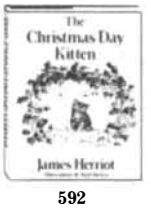


025
Pub. price $19.95

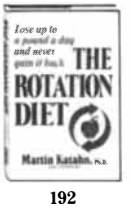


116
Pub. price $24.95

266
Pub. price $18.95



592
Pub. price $9.95

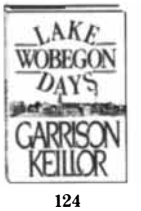


192
Pub. price $15.95

202
Pub. price $19.95

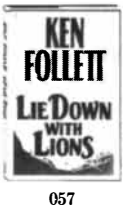
268
Pub. price $18.95

129
Pub. price $24.95



124
Pub. price $17.95

540
Pub. price $17.95

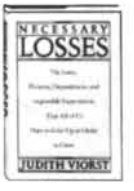


057
Pub. price $18.95

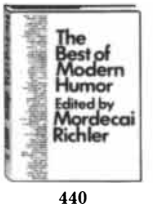
161
Pub. price $16.95

519
Pub. price $19.95

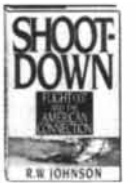
354
Pub. price $22.95



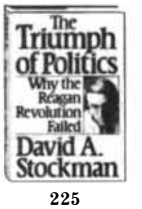
507
Pub. price $18.95



440
Pub. price $17.95

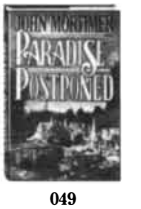


589
Pub. price $18.95

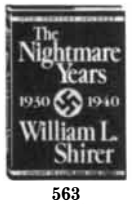


225
Pub. price $21.95

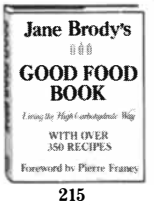
414
Pub. price $16.95



049
Pub. price $17.95



563
Pub. price $22.50



215
Pub. price $22.95

814
Pub. price $22.95

301
Pub. price $22.50

355
Pub. price $16.95

483
Pub. price $18.95

285
Pub. price $22.95

127
Pub. price $17.50

624
Pub. price $16.95

Choose any 2 books for $1 each now.

Then buy just one more book at any time during the next six months.

Facts About Membership. As a member you will receive the *Book-of-the-Month Club News*® 15 times a year (about every 3½ weeks). Every issue reviews a Selection and more than 125 other books, which are carefully chosen by our editors. If you want the Selection, do nothing. It will be shipped to you automatically. If you want one or more other books—or no book at all—indicate your decision on the Reply Form and return it by the specified date. *Return Privilege*: If the *News* is delayed and you receive the Selection without having had 10 days to notify us, you may return it for credit at our expense. *Cancellations*: Membership may be discontinued, either by you or by the Club, at any time after you have bought 1 additional book. Join today. With savings and choices like these, Book-of-the-Month Club is where book lovers belong.

Book-of-the-Month Club, Inc., P.O. Box 8803, Camp Hill, PA 17011-8803 A170-11

Please enroll me as a member of Book-of-the-Month Club and send me the 2 books I've listed below, billing me $1 each, plus shipping and handling charges. I agree to buy 1 more book during the next six months. A shipping and handling charge is added to each shipment.

Indicate by number the 2 books you want:

☐ ☐

Name _____ 5-50
(Please print plainly)

Address _____ Apt. _____

City _____

State _____ Zip _____

Prices generally higher in Canada.
All orders are subject to approval.

BOOK-OF-THE-MONTH CLUB®

COMPUTER RECREATIONS

Star Trek emerges from the underground to a place in the home-computer arcade

by A. K. Dewdney

In the early 1970's I often worked late at my university office in the hope of avoiding the interruptions of students. Unfortunately, just down the corridor in the computer-graphics research laboratory was the nightly gathering for a favorite student game variously called Star Trek or Space War. "Get him! Get him!" came the cries that echoed through my closed office door. "Watch out for the missiles!" The noise was just enough to disrupt my train of thought. Unable to beat Star Trek, I would usually join it. With a resolve to return to less frivolous matters as soon as the stretch had refocused my attention, I would stroll down to the lab to watch the action.

The game Star Trek is loosely based on the television series of the same name; the competition is a battle between the starship *Enterprise* and a Klingon battle cruiser. Aficionados of the television program will recall that *Enterprise* traveled where human beings had never ventured before. The exploration was done on behalf of a collection of cooperating races called the Federation, which was apparently dominated by people. Klingons were the hirsute rivals of the Federation for the domination of the Milky Way, if not of the entire universe.

Although Star Trek originally required a powerful graphics-research computer, it can now be programmed comfortably on a personal machine. The two starships orbit a central sun, launching missiles at each other and dodging the return fire. Both ships and missiles are subject to gravity, and orbital motion determines much of the action. Pilots without a reasonable feel for celestial mechanics incinerate their ships in the sun or unintentionally cross the deceptive trajectory of a missile's orbit. The side whose ship blows up first is the loser.

At one time versions of Star Trek had appeared at hundreds of universities and other institutions; the game was officially frowned on but secretly enjoyed. I find it ironic that years after those late-night interruptions I still recall the game with fondness. Star Trek was only one of the many games developed by students, and most such games soon found their way into commercial packages that were largely responsible for the revolution in home computers. Commercial versions of Star Trek have only recently been introduced, although arcade versions have been available for some time.

The version of Star Trek described here takes the reader back to the clandestine romance of the earliest computer games. Furthermore, it serves to introduce what I call the arcane subject of arcade programming. One goal is to keep the screen alive: the computations that create the action in the Star Trek world must be as fast and as simple as possible. I described a similar problem in July for programs that simulate flight. Luckily the Star Trek world is much simpler than the detailed, digital geography that must be presented to keyboard pilots. A second programming goal is to create a realistic gravitational environment, but here too readers have encountered the same problem in somewhat different guise. In January I introduced an armchair universe in which stars dance about in accordance with their mutual gravitational attraction. Gravity of a simpler kind warps the tracks of ships and missiles in Star Trek. Only the central sun exerts a perceptible force.

To engage in their orbital duel two players sit at the keyboard of a computer and press the keys assigned to control their ships. One player, perhaps the hairier one, takes command of the Klingon ship; the other guards the fortunes of the Federation. At first only the central sun and the two ships occupy the screen. The sun is a circle and the ships are icons with just enough detail for distinguishing friend from foe.

When the game begins, both ships are in free fall toward the sun. The players immediately turn their ships away from the fall line and fire their rocket engines to bring the ships into safe orbit. A ship that touches the sun instantly vaporizes and of course the game is lost.

As soon as the orbits are established, each player begins trying to eliminate his opponent. One extreme tactic is to lie in wait until the enemy ship passes nearby. A quick salvo may then finish it off. Another extreme is trickier: one can try to aim the shots from a position on the opposite side of the sun. On the screen bright points of light—called photon torpedos on the television program—fly outward from the firing vessel and burn their way around the sun in a gently curving array of menace. Unless the enemy ship is commanded by a pilot of extraordinary skill, the ship is destined to meet one of the missiles and explode in a burst of interstellar debris. The screen signals the event with a brief cloud of dots and announces either VICTORY FOR THE FEDERATION or VICTORY FOR THE KLINGON FORCES.

Because the action near the sun is so intense, cautious players prefer to orbit farther out. The main disadvantage of the strategy is the need to recharge a solar energy cell. The control of each ship depends on its energy cell. When the cell is spent, the ship must quickly move closer to the sun to replenish its supply of solar photons. In a distant orbit the photon stream is weak and the ship runs the risk of becoming a sitting duck.

There is another tricky feature to be aware of in high orbital flight: the battle space in Star Trek is "toroidal." In other words, if the ship moves too close to one edge of the screen, it will disappear there and reappear near the opposite edge.

Each ship has an infinite supply of missiles, but there can never be more than 10 of them in flight at the same time. The missiles obey the same laws of physics as the ships do. A missile lasts either until it strikes a ship (including its own) or until it runs out of fuel. So much for the game.

The program I call TREK is the most ambitious one presented to date in this department, not so much because of its complexity as because of its length. Some of the more standard routines can only be sketched in. Programming neophytes, particularly those already linked to advisers through the advisory network [see "Computer Recreations," SCIENTIFIC AMERICAN, December, 1985], may nonetheless try the project with some hope of success and considerable entertainment.

Artificial Intelligence:

One of the most significant developments in the history of computer technology since our invention of the integrated circuit.

TI's new generation of products and programs can dramatically enhance your productivity and put you ahead of your competition into the 21st century.

Knowledge T

Texas Instruments brings the practical

"AI will be a significant force in the worldwide competition for productivity leadership . . . and a key enabler for the new billion-dollar markets of the future. TI can help you lead the way."

— Dr. George Heilmeier, Senior Vice President and Chief Technical Officer, Texas Instruments.

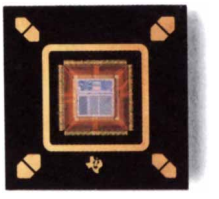
Artificial Intelligence is hot news. New jargon, new ventures and new promises abound. Some struggle with definitions and debate its import for mankind . . . while uncertainty prevails over action.

echnologies:

pplications of AI to your business. Now.

But at Texas Instruments, the vision is clear. And the vision has become reality: we are putting AI to work under a company-wide strategy we call "Knowledge Technologies*." These are the real products, programs and tools required to aggressively pursue the potential of AI through practical applications — throughout our own company and for our customers.

In 1978, TI launched one of the world's largest commitments to Artificial Intelligence. We saw it then and see it now as a driving force in the next generation of computing technology and thus a cornerstone to future competitive productivity. Today, we offer the broadest array of products, tools and programs available to help you realize its bright potential.



The Explorer MegaChip LISP processor† can replace hundreds of chips in today's LISP machines.

What data has been to the first generation of computing, knowledge is to the second.

Knowledge is the capital of the future. If you manage it as the strategic resource it is, its impact on productivity can have powerful economic repercussions in the decades ahead.

Coupled with Personal Consultant Plus software, TI's Business-Pro* AT-class PC is a powerful development and delivery tool for expert systems.*

Intelligence on a chip.

At the heart of Knowledge Technologies lies the development of a pioneering semiconductor chip. We developed it under contract to the U.S. Government for use in aerospace and defense, but its impact will be felt in all areas. One of its first commercial applications will be to enhance the power and performance of the TI Explorer* computer, already one of the world's most advanced AI development tools. Also, it will add new members to the Explorer family of products.

The Explorer system is an advanced work station ideally suited for the rapid development and delivery of AI software and for conventional software.

This advanced chip, called the Explorer MegaChip* LISP processor, promises to greatly extend the range of AI applications. Intelligent machines will be easier to use, easier to obtain and easier to afford. On the horizon are machines that will communicate in a language everyone understands; speak, hear and see; solve complex problems requiring inference and deduction; and help make complicated and subjective decisions.

▶ *See back page for more information.*

Knowledge Technologies as a competitive edge.

Retaining their leadership position in the late '80s and beyond was the driving force behind Campbell Soup's search for new technology. They discovered a very powerful tool called AI. And TI's Knowledge Technologies put it to work. They saw it as a competitive edge, an advantage they could build on to ensure their future success.

In Campbell's case, their first application of AI yielded an expert system that incorporated 44 years of one man's troubleshooting know-how and delivered it to maintenance workers in Campbell Soup plants across the nation. Instead of relying on one man's expertise, the system can diagnose problems and recommend actions through dialogues on a standard PC. The system has worked so well that three others are in development, with more planned for the future.

Expert systems and knowledge engineers.

The expert system mentioned above was developed in seven months by Campbell's "expert" and a TI knowledge engineer using Personal Consultant* software, which provides an efficient way to build expert systems. In order to answer the growing demand for knowledge engineers, TI has created a Knowledge Engineering Department with professionals trained to identify appropriate expert system applications, organize knowledge into rule-based structures and create prototype expert systems.

Expert systems promise to be invaluable productivity tools in manufacturing, education, medicine, finance and law — almost every field requiring human expertise. Applications will include prediction, diagnosis, design, planning, alarm analysis, monitoring, simulation, repair, instruction, interpretation and control.

Start with TI. Stay with TI.

We have tools that can help get you started, no matter where your company is on the AI learning curve. We have courses and programs that can take the mystery out of AI and help you to see where it fits into your company. We can help you build expert system prototypes so you can safeguard and multiply your company's human inventory of knowledge and experience.

TI's AI Education Center in Austin, Texas, is equipped to handle your AI education needs.

Or we can teach you how to use our tools to build your own expert systems. Or how to enhance your own software productivity by solving software design problems and functionally testing in prototype form.

Our portfolio of tools includes a broad range of software, hardware, training and support. And we have ongoing programs designed to keep you current with the technology. It doesn't take a huge investment, either. For under $1,000 for Personal Consultant software, you can start building expert systems on an industry-standard PC.

Take your first step into the future.

It's important to start now. And because we offer the widest array of products, tools and programs, the best way is to start with us. We can help you identify your company's projects where AI can be applied, where your people can gain experience quickly, and where we can help you prepare for the future.

Write us on your letterhead. Tell us whom you'd like us to talk to in your organization, and we'll take it from there. We'll get you started in AI. We think once you start with TI, you'll stay with TI.

"Texas Instruments has made a major, long-term commitment to Artificial Intelligence. We have applied AI aggressively to our own productivity and products. And now we are helping our customers realize AI's potential through a wide range of products, services and support."

Jerry R. Junkins

— Jerry R. Junkins
President and Chief Executive Officer

Write to: Data Systems Group, Office of the President, Mail Station 2068, Texas Instruments, P.O. Box 2909, Austin, Texas 78769.

TEXAS INSTRUMENTS

31597

* Knowledge Technologies, Business-Pro, Explorer, MegaChip, Personal Consultant, and Personal Consultant Plus are trademarks of Texas Instruments Incorporated.

†MegaChip IC size — .3937" x .3937"; actual package size — 1.75" x 1.75".

TREK cycles through six major sections of code as long as both ships are operational:

Read the keys.
Update ship and missile positions.
Check for contacts.
Update energy of ships.
Manage missiles.
Display.

With the possible exception of the first section, many readers will find that TREK is relatively straightforward to write. Reading keys will be new to some, but the facility is indispensable for arcade programming. Most high-level languages have statements that enable a program to test whether a particular key has been pressed.

For example, in the language Microsoft BASIC the relevant command is

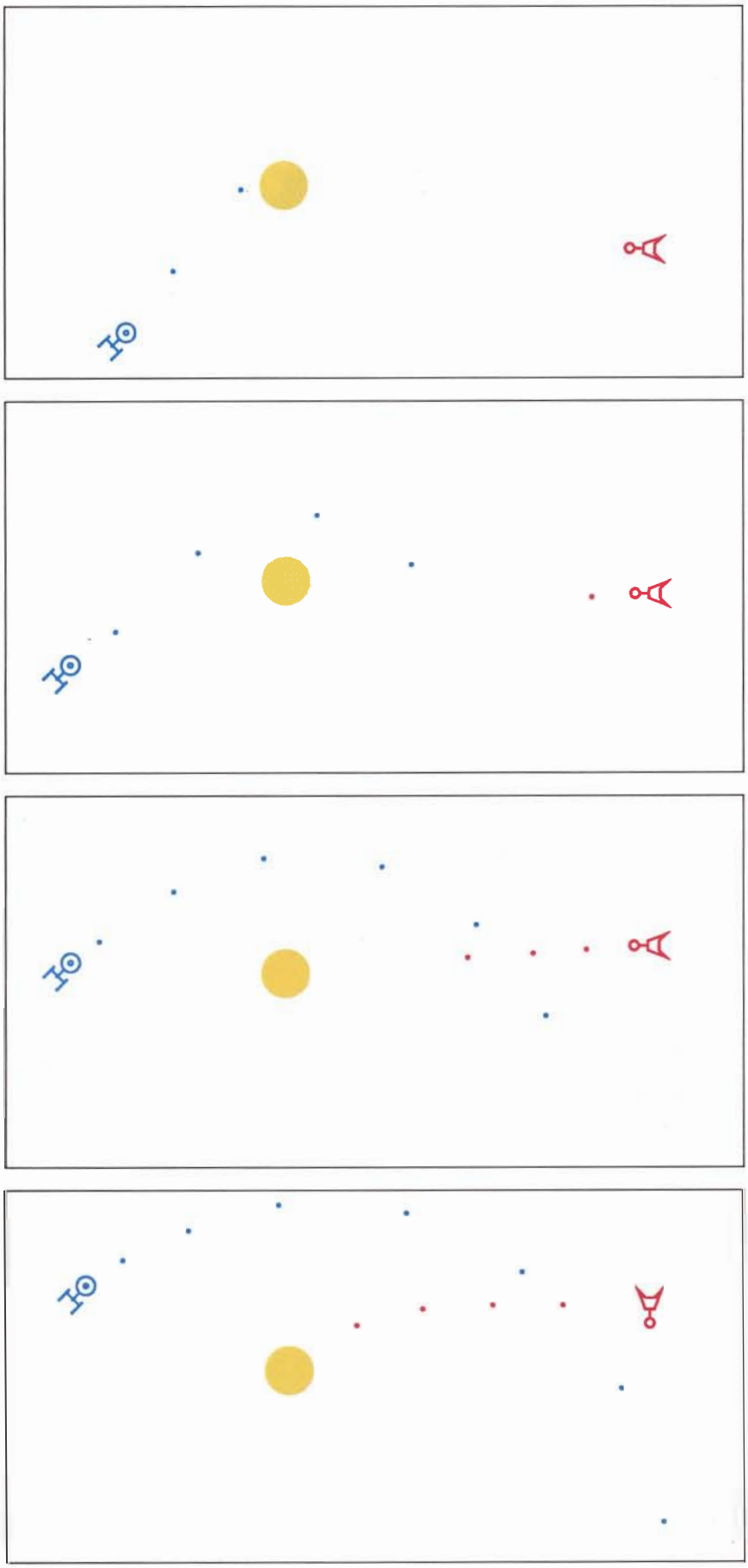
On key(k) gosub n.

When the "On key" command is executed, the program checks at the beginning of every subsequent command to determine whether or not key k was pressed. If it was, the program branches to the command at line n. At line n a subroutine begins whose purpose is to record the pressing of key k, generally by assigning some value to a "flag" variable. There are 14 specific keys on the keyboard of the IBM PC that can be checked in this way: the 10 function keys and the four cursor-control keys. One must assign four keys to each player—say function keys $F1$, $F2$, $F3$ and $F4$ to the Federation forces and the four cursor keys to the Klingon forces. The numbers for this assignment are respectively 1 through 4 and 11 through 14. A manual is indispensable here.

For each side the first of the four keys controls thrust, the next two control direction and the last one controls the firing of missiles. The thrust key simply imparts a fixed thrust, at full throttle, for one program cycle. Each press of a direction key turns the corresponding starship through an angle of 10 degrees. A tap on the missile key fires a single missile. Before playing the game it is useful to stick small square labels on the controlling keys. Appropriate symbols marked on the labels can then remind the combatants which key does what.

The program TREK must include an "On key" command for each of the eight designated controlling keys. "On key" is not executed unless it is preceded by the command "Key(k) on."

All the subroutines to which TREK branches from "On key" commands are simple and essentially the same. Each subroutine is made up of two in-



Enterprise *and the Klingon battle cruiser exchange missiles*

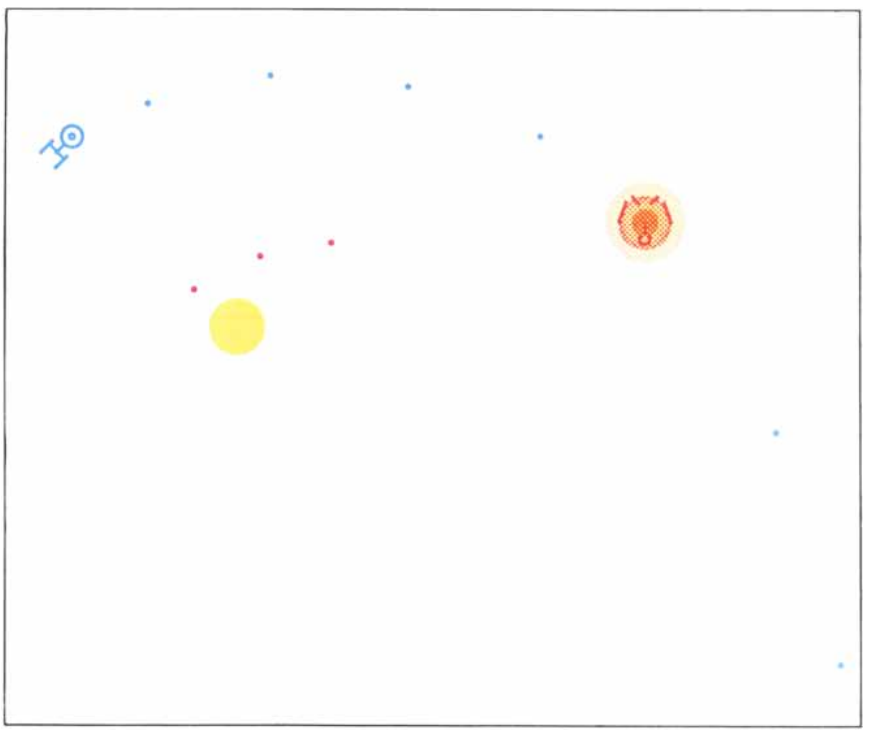
structions. The first instruction sets a flag variable to 1 for later consultation by the program; the second causes a return of the program execution to the line number of the "On key" command that invoked the subroutine. The flag variables for the control of *Enterprise* might as well be called *fdgo, fdrt, fdlt* and *fdfr*. They stand for "Federation go" (turn on thruster), "Federation right" (turn clockwise), "Federation left" (turn counterclockwise) and "Federation fire" (shoot a missile). Similarly, the variables *kngo, knrt, knlt* and *knfr* record the thrust, direction and missile firing of the Klingon battle cruiser.

When a flag variable is set to 1 within a subroutine, it triggers a change in one of the starships. For example, when *fdgo* is 1, the position-updating segment of the program adds a small acceleration (between 2 and 5, according to taste) to the current acceleration of *Enterprise*. TREK must then reset the flag to 0.

Updating the positions of two starships and a handful of missiles is much easier than managing an equal number of massive stars. The combined mass of warring hardware is trivial compared with the mass of the central sun, and so the mutual gravitational attractions of ships and missiles are assumed to be zero. Given the distance of each object from the sun, TREK simply calculates the acceleration of the object caused by the centrally directed solar gravity, updates the velocity of the object and finally revises its position.

Even for such a conceptually simple calculation there is substantial computational overhead. Sums, products and square roots are needed to carry out each calculation. To avoid slowing down the game with excessive arithmetic, TREK consults a table; for every possible distance from the sun an array called *force* gives the predetermined acceleration experienced by an object [*see upper illustration on opposite page*]. Since the Star Trek universe is toroidal, the new position derived from the acceleration found in the *force* table must be calculated modulo the horizontal or vertical distance across the rectangular display.

Two arrays give the current velocity and position for the two starships and as many as 20 missiles; the arrays are called *vel* and *pos*. Each array has two columns and 22 rows. The first two rows hold starship data and the next 20 are devoted to missiles. Thus in the first row of *vel* the two entries *vel*(1,1) and *vel*(1,2) are respectively the velocities in the *x* (horizontal) direction and the *y* (vertical) direction of *Enterprise*. Similarly, *vel*(2,1) and *vel*(2,2) give the two mutually perpendicular velocity components of the Klingon battle cruiser. A special variable called *misnum* enables TREK to keep track of the number of missiles currently in flight. *Misnum* ranges from 2 (no missiles) through 22 (20 missiles).



The Klingon ship is not able to dodge the barrage

By lumping starships and missiles together in the arrays one creates a shorter and slightly more efficient program. Only one loop is needed to update both positions and velocities. Before the update, however, two variables are needed to keep track of the orientations of the two spacecraft: *fdor* and *knor*. Their values are expressed in degrees, where the angle 0 indicates a ship is pointing east, 90 indicates north and so on. Every time a player presses either direction key, one of the two variables is increased or decreased by 10 degrees as appropriate.

The loop for updating positions is indexed by the letter *i,* which ranges from 1 through 22. For each value of *i* the program calculates the acceleration: the current values of the *x* and *y* coordinates of the *i*th row of *pos* are squared and added together. TREK then finds the square root of the sum, and that number is truncated to the greatest integer less than or equal to it. The truncated integer square root is the approximate distance from the sun to the position of the *i*th object. That integer is then taken as an index to the acceleration table, where the solar attraction can be looked up.

For the two starships the components of solar acceleration must then be added to the components of thrust. If either flag variable *fdgo* or *kngo* has been set equal to 1 (in other words, if the thrust key has been pressed) and if neither starship is out of fuel, TREK must multiply the thrust constant by its horizontal and vertical components. For my monitor I take the thrust constant to be 3, which gives reasonable maneuverability to the spacecraft. For *Enterprise* the horizontal component is then three times the cosine of *fdor;* the vertical component is three times the sine of *fdor*. For the Klingon ship the thrust components are derived from the angle *knor*. As soon as the calculation is done TREK resets both *fdgo* and *kngo* to 0; the throttle is turned off until the next time the thrust key is pressed.

When the index *i* becomes greater than 2, the calculation of thrust can be bypassed because missiles have no thrust. The rest of the updating loop is devoted to calculating new velocities and positions for the moving objects on the screen. For each object the numerical magnitude of the acceleration is added to that of the velocity, and the magnitude of the velocity is added to that of the position. Such a simplistic calculation is made possible by adjusting the thrust and solar attraction to reflect a system of units that assumes the passage of one unit of time for each program cycle.

To check for contacts among the various objects the program must first determine for each ship whether the ship lies on or within the boundary of the sun. Since TREK has already calculated the updated distance between each ship and the sun, the program needs only to compare that distance with the solar radius, say 10 units. If either ship has collided with the sun, TREK responds with an appropriate screen message, such as KLINGON VAPORIZED; the program then branches to its display segment.

A second check for contact must determine for each missile whether the missile lies within a certain small distance of either ship. Here TREK uses a simple but effective shortcut: it finds the difference between the x coordinates of the missile and a ship, and it does the same for the two y coordinates. Finally it adds the two differences; the process avoids both squaring and taking square roots, and the result is nearly as good as the usual distance calculation. If the sum of the two differences is less than, say, 4, the program scores a hit. A message appears on the screen, such as ENTERPRISE HIT BY A MISSILE. KLINGON WINS. A single loop carries out the test for each missile. Its index starts at 3 and ends at *misnum,* the number of missiles in the sky plus 2.

To update the energy levels of the ships the program divides the solar acceleration obtained from the table by 60. Since the acceleration increases as the ship moves closer to the sun, such a ship can receive a more concentrated stream of energetic solar photons. The energy is then added to a fuel variable called *fdfl* or *knfl,* depending on which starship is involved. Each of these variables is decreased by .1 when thrust is applied; the decrements are made in the position-updating segment of TREK. A starship is considered out of fuel if it has no more than one unit of solar energy left in its tank. The tank begins life with 10 units of fuel.

Missile management requires an array called *time,* which stores the number of program cycles in the life of each missile. When a cycle count reaches 25, the corresponding missile is removed from the array *pos,* the count is reset to 0 and *misnum* is decreased by 1. The missile can be removed from *pos* in one of two ways. The first method is easier to program but may slow down the game. TREK runs through the array from the index value at which the missile is removed and decreases the row number of each entry by 1. Thus the last entry to be shifted lies at index *misnum*. It is moved to row number *misnum*-1. The same operation is carried out on the arrays *vel* and *time.*

A faster technique takes advantage of the observation that the oldest missiles have the smallest indexes; they were the first missiles added to the list. One can therefore keep track of the missiles without shifting their indexes; missiles whose age has reached 25 program cycles must all be found at the beginning of a contiguous group of missiles in each array, and so only they must be removed. Similarly, new missiles are always added at the end of the contiguous group.

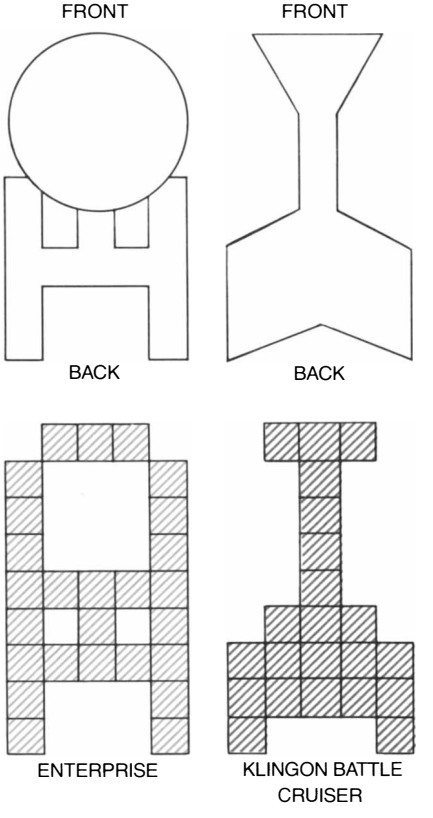
Introduce two new variables called *old* and *new,* which serve as pointers to the oldest and newest missiles in each array. As missiles are removed and added, only the values of *old* and *new* must be changed. One can then apply modular arithmetic to keep the contiguous group of missiles cycling around in each array. When a new missile is to be added at index value 23, TREK reduces the index modulo 23 to 0 and then adds 3 to avoid replacing one of the starship coordinates by a missile coordinate. The variables *old* and *new* undergo the same process. Such a data structure is called a circular queue. If this arcade trick is used, the position-updating segment of the program must be modified: split each single loop into two smaller loops, one for ships and the other for missiles.

When a player presses a key to fire a missile, TREK first checks a count of missiles currently activated by that side. If the count is less than 10, TREK consults the values of the flag variables *fdfr* and *knfr*. If *fdfr* is 1, for example, the program adds 1 to the missile count for the Federation, increases *misnum* by 1 and then loads the position and velocity coordinates of *Enterprise* into the appropriate slots of *pos* and *vel*. In the process the program should add four units to the position coordinates and two units to the velocity coordinates; in both cases the addition is made along the same direction as the ship is currently moving. For example, the horizontal position coordinate of a missile fired by *Enterprise* is four times the cosine of *fdor* plus the horizontal position coordinate of *Enterprise;* the vertical coordinate increases by four times the sine of *fdor*. The initial position of the missile is thereby kept clear of the ship, preventing immediate destruction by the ship's own photon torpedo. The same operation applied to the missile's velocity coordinates reflects a relative launch speed of two units per cycle: a missile travels two units per cycle faster than the ship that launches it.

The last major section of TREK displays the sun, two ships and whatever missiles are active at a given time. The program draws a circle of radius 10 in the center of the screen and then works its way through the array *pos*. The Federation and Klingon ships are represented by icons. One icon is essential-

DISTANCE	FORCE
10	8.000
11	6.612
12	5.556
. . .	
	$\text{FORCE} = \frac{800}{(\text{DISTANCE})^2}$
. . .	
178	.025
179	.025
180	.025

Sample force table and its formula



Icons **(top)** *and their pixel versions*

ly a circle that recalls the famed discoidal *Enterprise* with its twin engine booms. The Klingon icon is angular. Readers are free to attempt any miniature variation on these ships as long as two things are reasonably clear at a glance: which ship is which and where each one is headed. In drawing the icon for either spacecraft, TREK calls on a list of display points that must be translated and rotated to reflect the position and orientation of each object. For these operations the program consults the array *pos* and the variables *fdor* and *knor*.

Missiles are simpler. The display program draws each missile as a point within a single loop, consulting *pos* as it goes along.

TREK is subject to one disadvantage of the display screens now in popular use. Such screens operate in storage mode: an object drawn on the screen remains there. To avoid a confusing welter of remnant ships and missiles TREK must draw each object twice. It first draws the object in its old position in black. Then it redraws the object in its new position in the normal color.

I shall leave the details of initializing the program to the Trekkers who attempt it. In spite of arcade programming, some of you may find the game too slow; you may be tempted to call it Star Truck. For better performance try compiling your program, or impose arms limitations on the number of missiles allotted to each side.

I have described only the bare-bones version of Star Trek. Fancier but private editions have been built and they continue to propagate; games have appeared that allow three or more spacecraft, laser guns, color graphics and status displays. I must thank Jonathan N. Groff of Clearwater, Fla., for reminding me of this underground classic and for introducing me to a version of the game that includes an automated Klingon. Earthlings representing the Federation are continually wiped out by Groff's program.

For the rest, I invite readers who manage to create a working version of TREK to write to me with an account of their own strategies for arcade acceleration. I shall include the best results in a later column.

The vast majority of the letters responding to the discussion of computer magic in the August edition of this department replied to a question answered in the September edition: How did the magician, with access only to a hand calculator, read the subject's mind? The answer depended on the fact that any number made up of digits that are colinear on the calculator key pad is divisible by 3. It is not my intention to give a full explanation of the trick once again, but I do want to distinguish between two kinds of questions that are regularly asked in this department.

An intellectual encounter of the first kind usually involves a purely mental poser that will be answered in the very next issue. A warning that the answer is immediately to follow accompanies such a question: That answer must be at the press long before the first hopeful letters arrive.

An intellectual encounter of the second kind may depend on mind or matter. A computer and some thought are normal prerequisites for solving this kind of question, and readers are regularly invited to respond to it. Here there is time to tinker. A response mailed in within two or three weeks of the magazine's appearance has a good chance of crossing my desk before my summary of the replies must go to press. The summary generally appears three months after the article that prompts it.

A twist to the calculator trick was suggested by Carl Fulves of Teaneck, N.J. As before, the magician asks the volunteer to select the two colinear numbers and to multiply them using the calculator. In my discussion I had suggested that the volunteer withhold one of the numbers from the product; the magician could then "guess" it by casting out 9's. Somewhat more impressive is the addition of some canine (or feline) misdirection. The volunteer withholds nothing and instead adds the age of a pet. If the age is not greater than nine, the magician can easily recapture it: "Madame, I divine Tinkerbelle's age to be five."

Russell L. Mullennix, an engineer with the Georgia Power Company in Valdosta, Ga., told me of a brief but instructive adventure with PREDICTABIT, the program that attempts to predict what bit a volunteer will think of next. After a reasonably long stream of 0's and 1's, the program's predictions should improve. They might do so because of unconscious patterns in the volunteer's efforts to be random. Mullennix had modest success with a colleague, scoring percentages in the high 50's. When his colleague caught onto the program's probable modus operandi, however, he began to select bits based on the digits of pi. The program's prowess dropped to a dismal 43 percent. Readers cannot say they were not warned. Subjects must either be carefully selected for credulity or be kept off balance by a patter that prevents them from thinking about what they are doing.

Radar warning breakthrough #4 is now available from the same engineers who made #1, #2, and #3

Bad news for radar detectors. The FCC (Federal Communications Commission) has cleared the Rashid VRSS for operation on K band.

What's a Rashid VRSS?

The Rashid VRSS is a collision warning system using a radar beam to scan the vehicles path, much as a blind person uses a cane. It may reduce accidents, which is very good news.*

Now for the bad news

Unfortunately, the Rashid transmits on K band, which is one of the two frequencies assigned to traffic radar. Rashid speaks a radar detector's language, you might say, and it can set off detectors over a mile away.

Faced with this problem, we could hope Rashid installations will be few. Or we could invent a solution.

Opportunity knocking

Actually, the choice was easier than it sounds, because our engineers are in the habit of inventing remarkable solutions. In fact, in the history of radar detection, only three advancements have qualified as genuine breakthroughs, and all three came from our engineers.

Back in 1978, they were first to adapt dual-band superheterodyne technology to the problem of traffic radar. The result was ESCORT, now legendary for its performance.

In 1983, when a deluge of cheap imported detectors was found to be transmitting on radar frequency, our engineers came through again, this time with ST/O/P™, a sophisticated circuit that could weed out these phony signals before they triggered an alarm.

Then in 1984, using SMDs (Surface Mounted Devices), micro-electronics originally intended for satellites, these same engineers designed the smallest detector ever. The result was PASSPORT, renowned for its convenience.

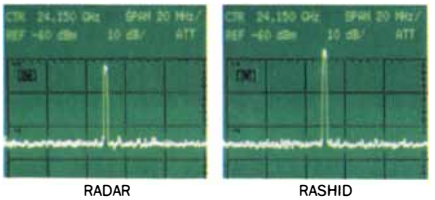
*For more information on Rashid VRSS collision warning system, see *Popular Science*, January 1986.

They said it couldn't be done

Now we're introducing breakthrough number four. In their cleverest innovation yet, our engineers have found a way to distinguish Rashid from all other K band signals. It's the electronic equivalent of finding the needle in a haystack. The AFR™ (Alternating Frequency Rejection) circuit isolates and neutralizes all Rashid signals, yet leaves the radar detection capability undiminished for your protection.

No waiting for the good stuff

When testing proved that AFR was 100 percent effective, we immediately incorporated it into ESCORT and PASSPORT. Our policy is to make running changes—not model changes—whenever a refinement is ready. That way our customers always get the latest science.



RADAR RASHID

Figure 1: A digital spectrum analyzer scanning the entire width of K band can't see the difference between radar and Rashid.

AFR is fully automatic. There are no extra switches or lights. Nothing for you to bother about. The Rashid problem simply goes away.

Last year *Road & Track* called us "**the industry leader in detector technology.**" We intend to keep earning our accolades.

We make shopping easy too

Call us toll free. We'll answer all your questions. If you decide to buy, we'll ship the next business day at our expense. For $6.00 extra, Federal Express will deliver to you within two business days of shipment.

If you're not completely satisfied within 30 days, return your purchase. We'll refund *all* your money, including return postage, no questions asked.

We specialize in breakthroughs. Can we make one for you?

Order Today

TOLL FREE...800-543-1608

(Phone M-F 8-11, Sat 9-5:30, Sun 10-5 EST)

VISA MasterCard

By mail send to address below. All orders processed immediately. Prices slightly higher for Canadian shipments.

PASSPORT®
RADAR·RECEIVER

Pocket-Size Radar Protection $295
(Ohio res. add $16.23 tax)

ESCORT®
RADAR WARNING RECEIVER

The Classic of Radar Warning $245
(Ohio Res. add $13.48 tax)

Cincinnati Microwave
Department 916N
One Microwave Plaza
Cincinnati, Ohio 45296-0100

Cincinnati Microwave is committed to constant advancement in radar warning technology. But we don't believe in planned obsolescence. Therefore, we are working out a plan to offer upgrades for most pre-AFR models (PASSPORTs under number 550,000, and ESCORTs from 200,000 through 1,199,999). Since non-AFR units will require extensive modifications, new procedures and facilities are being developed to ensure a quality conversion. For details and costs of this retrofitting program, see our ad in the January issue of this magazine.

BOOKS

The birth of zero, the talking-chimp debate, the stuff of fleas—and a rebuttal to Thurow

by Philip Morrison

FROM ONE TO ZERO: A UNIVERSAL HISTORY OF NUMBERS, by Georges Ifrah. Translated by Lowell Bair. Viking Penguin Inc. ($35). The accountants have provided that most potent of contemporary finalities, the bottom line. Now it appears that they produced the top line too. The first writing, according to Georges Ifrah, seems to have been "invented by accountants who had to deal with economic operations much too complex and varied to be entrusted to a single memory."

The time was 50 centuries back; the places were Sumerian Uruk on the old Euphrates and Elamite Susa east of the Tigris. The first documents are clay tablets, marked while they were soft with a simple pointed tool (not yet the sedge used for cuneiform). The entries show pictographs for date palms and mutton legs. Adjoining them are coded marks for numbers. There is a precursor to this full-fledged account book: a couple of centuries earlier we find clay envelopes that resemble a tennis ball in size and form. They enclose a set of small clay tokens, almost certainly themselves the descendants of pebble counters. (An X-ray photograph of such a specimen at the Louvre is redrawn here.) The tokens are numerical symbols; a little sphere, say, means 10 units. An impression from a cylinder seal on the outside authenticates the secured record.

Within the past decade various French expeditions and scholars have uncovered level by level a scheme that was steadily elaborated in the course of a few centuries. First, the envelope becomes merely a legal record; its numerical content is marked again on the outside so that it can be read without destroying it: the stylus marks match the enclosed tokens. Next the envelope evolves into marked, solid tablets of account, again validated by cylinder-seal impressions. The fourth step at Susa shows minor neatening. In Uruk at that same time new tablets appear, between 3200 and 3100 B.C., bearing engraved marks. These marks sometimes box together the numerical symbols based on the tokens with new signs that code for the commodities being counted. The symbols are as realistic as a sketch of a bird, or as abstract as a crossed circle representing a sheep. Writing has certainly appeared, although not as yet the richness of a real written language. Accountants tend to be frugal.

A delightful illustration shows the basis for the identification of Sumerian numerals: the scribes followed the helpful practice of putting totals on the reverse of many tablets. The example shown is a formal accounting for 15 bags of barley, 30 bags of wheat and so on. The total is written on the overleaf (overclay?) as 145 bags in all. Given many tablets, the checking method allows unique decipherment.

This interesting volume by a gifted and enthusiastic amateur scholar (a polyglot French teacher of mathematics) offers almost 30 chapters that range over the entire story of number systems, both written and unwritten. Ifrah draws all the many figures in his "archeology of numbers" with his own hand; the unity and clarity gained are a pleasure for the reader. There is a price: the sense of evidence is necessarily diluted by all those figures plainly written in the same style, whether they originated as Chinese, Hebrew, Roman or Zuñi.

Most of the story can be seen as a tension between the early iterative and the later structural. One-to-one correspondence lies at the taproot, but that seems no very new insight for our species. There are good accounts of the mapping of numbers on body parts, not merely the 10 fingers to which all forms of numeration except for machine language still bear homage, but systems such as those used in the Torres Strait, where ankles, knees, hips and more gave a visual representation that ran up to 33. A table of the first 10 number words in a couple of dozen Indo-European tongues, from Sanskrit *dvi* and *tri* to Icelandic *tveir* and *prir,* bears out their kinship. The words have lost all meaning but the digital (that trendy word has kept both meanings); were our number words also once "names of parts of the body"?

The number base is a brilliant recursive early invention. The role of anatomy in its origin is clear. The pedigree of base 60, however, is obscure. We use it for telling time and in circular-arc numeration, a practice we can ascribe to the Babylonians and to the Sumerians before them. A little eased by an auxiliary use of 10, it saved them the labor of inscribing symbols. If 10 was *u* and 60 was *gesh,* it was nice if not fully consistent to call 600 *gesh-u.*

Why 60? In a fine page Ifrah collects the theories, by now antique. Theon, a Hellenic author, was the first to offer one answer, known to generations of the mathematical: 60 is uniquely full of divisors. Perhaps two peoples, one using the base 6 and one the base 10, came together—an unfounded conjecture made in 1904. One subtle argument began with the Babylonian double hour, a twelfth of the full day, spanning an arc 60 times the angular diameter of the sun.

Cantor himself, the founder of set theory, held that the beginning was the rough day count of the year, 360; then the simplicity of the hexagonal division of the circle, the chord equal to the radius, further privileged 60. Cantor gave this up when critics objected that neither astronomy nor geometry could explain a numeration system. Once one learns that Chinese reckoners long ago divided all circles not into 360 degrees but into 365¼ degrees, he is led to think Cantor ought to have held firm. Perhaps in the oldest times few people other than astronomers needed to compute with large numbers. The early tablets already use the standard Sumerian system with its mixed reliance on 10 and 60.

To the Babylonian sexigesimal computers we owe both the first place-value system of notation and the first use of zero. Their entries are strictly positional in the base 60, although their cuneiform numerals are formed by symbols for 10 and 1 combined in a simple way (thus 32 is made of three marks for 10 and two for 1). The notation for 610 requires a crescent for 10 in the 60's place and then another crescent for 10 in the units place. But now the scribe was worried: that might look like 20, two crescents side by side. He spaced his marks carefully. After 1,000 years or so, practice had moved ahead. The sign for a separation, used in many prose texts for such purposes as marking the end of one language and the start of another in bilingual documents, was drafted to fill the necessary empty space. In the Babylonian

mathematical documents that device appears only when the zero is medial, never at the start or end of a number. The astronomers went further; they used it wherever necessary.

It remains uncertain that the sign could be read as a true abstract null, say the result of 20 minus 20. Instead the authors of some texts resort to a more concrete statement, for example "the grain is exhausted." Such was the dawn of a written zero, an event that probably preceded the first records we have of the zero, which date to the fourth century B.C.

The elements of our 10-digit system are called Arabic numerals, for thence they came to the West; yet they are almost surely southern Indian. They can be seen here in a variety of Indian documents dating from the sixth to the ninth century A.D. The decimal place-value notation, with its zero explicit and free, most of its digital forms swift and abstract, is there in full consistent use; even a few of the signs are familiar, although not all 10 are recognizable. That numeration is the "only real universal language" of our times. The original forms were entirely Indian; the system itself might have owed something to old Babylon. We do not know; it is most likely that the entire scheme was an Indian invention whose antecedents are the place notation and the place holder in a counting board. The treatment here of the many documentary and interpretive issues involved is particularly full. The idea of a mark for nothing touched the imagination of Europe; it seemed almost magical—in a way it is—and the word recalls that sense. The word comes from the Arabic for the void, ṣifr, after the Sanskrit with the same meaning. In English we hear in the word cipher the general notion of number and a hint of the secret or mysterious as well.

The volume is a full and clear account. It is at its best close to the documentary record; it loses a good deal of authority when it discusses the origins of counting as described in ethnographic and psychological literature, where secondary sources and speculation perforce take a larger role. There is plenty of fine reading here as well as quite a few sign and number puzzles. Although they are unraveled, they should still challenge readers who have a bent for such tasks. We miss an index, and the inviting citations, particularly for the wealth of figures, are unfortunately incomplete, a style not unusual in France.

GAVAGAI! OR THE FUTURE HISTORY OF THE ANIMAL LANGUAGE CONTROVERSY, by David Premack. The MIT Press ($12.50). Original and engaging, this small book delights by form and content alike, constructing a personal work of art in the act of making an analytic response within a protracted academic controversy. With good luck this form will spread to other fields, although it will never be easy to find research workers at once so good-humored, informative, telling and deep as David Premack, a California experimenter who has spent 20 years probing the mind of an ape.

Only an introduction (and one table of results) is spent on looking back in a rather generous mood at the hot controversy itself. Premack and his colleagues saw a "discursive representational competence" in the chimpanzees they taught to choose and aptly arrange so many plastic word tokens. The reports of these results provoked heavy skepticism. The animals were held by the critics to be mere imitative followers of hidden social cues that were given unwittingly by the experimenter. The response here is methodologically as powerful as it is gently put. Controls for such cues never did change the results. Moreover, deliberately introduced, conspicuous social cues were often ignored. No critic ever pointed out the inadequacy of specific controls, or made the claim that such controls are impossible. "Instead, polemic raged." Human uniqueness was seen to be at stake; the agenda was external.

The rest of the book looks forward, to research on other species, to newer results with apes, to an informal and skeptical review of the deep foundations of interspecific linguistics, often put in the form of thought experiments, instructive and charming to the inexpert reader. One of these centers on the code word that forms the enigmatic title.

"Gavagai" is a word invented a generation ago for a thought experiment of his own by the philosopher-logician W. V. Quine. It was the meaningful utterance of a total stranger; a linguist interested in translation of the radically different tongue assumes it is a word and tries to grasp its meaning by varying the stimulus while eliciting the speaker's assent or dissent. Even if the utterance always comes when a rabbit is at hand and is never elicited by small dogs or white frogs, does it really mean "rabbit"? Perhaps it refers in a quite unfamiliar way to mere temporal segments of rabbits, or even to all and sundry parts of rabbits, or to the fusion of all rabbits. For those are all present whenever any single rabbit is, and absent when nothing rabbity is around. Quine hopes thus to demonstrate that nonverbal tests cannot show the nuance of meaning.

In fact, actual studies of nonverbal infants already have gone far enough to test to a degree their idea of object constancy. In such a way the gavagai concept might be tested as well, say by a delightful videotape in which rabbit figures fly through space to come apart and reassemble in new ways. Such a tape will surely amuse most adults, producing smiles, even laughter; in children it may even stir a bit of fear. For we humans see the live rabbit as an enduring whole. But if gavagai is the common word for rabbit parts, not wholes, such a cartoon becomes too expected to be funny.

The world is big, and it can be partitioned in many ways. In science as a whole no inference easily becomes certain, let alone in translation from the who-knows-what. Psychology and translation are difficult, to be sure, without direct information on internal states. Is that difference central? Premack thinks not; he elaborates a variety of ingenious ways to interrogate without words. A quantitative scale might arise here someday, but the reader is apt to grant that the power of nonlinguistic tests ought not to be limited a priori without more knowledge than we have today.

Is recursion the key to human language? Imagine a cerebral mutation that allowed the idea of discrete infinity, a concept that links the idea of number to the innate human ability to make word strings unendingly fresh. (The notion is quoted from an unpublished comment by the invariably stimulating Noam Chomsky.) Chimpanzees have had an opportunity to show their mastery of recursion; so far they have not shown it. One experiment cited is a pleasure: pieces of fruit were hidden in a field while the apes watched. Then they were allowed in to retrieve the invisible fruit (from memory, of course). Their paths were recorded. They were economical, with little retracing of steps. But alas, their paths could be generated by a merely iterative algorithm: the nearest neighbor should be visited next, applied over and over. Recursion did not enter.

If one theme shapes this fascinating and candid account of ideas and opinions, it is that language, spoken or symbolic, is but the outward representation of a rich inward multiplicity of states. The human specialization appears to be social communication. Yet there are many other social dispositions, as Premack calls them, that may link humans even more deeply than our communicative specialty does.

"Pedagogy, aesthetics, cognition, social attribution, and consciousness, like language, are some of the major human specializations. To get on with

an understanding of our species, we shall have to relinquish our infatuation with language and stop reading and writing papers like the present one." People who write history usually conclude that civilization begins with writing; people who study language deeply often conclude that human intelligence is basically linguistic. A compelling account of the nature of human nature is still to come.

A Colour Atlas of Insect Tissues Via the Flea, by Miriam Rothschild, Yosef Schlein and Susumo Ito. A Wolfe Science Book, published in the U.S. in association with Methuen, Inc., 29 West 35 Street, New York, N.Y. 10001 ($79.95). The proper study of mankind may not be fleas, but these three enchanted biologists will be slow to concede that. Most of the 260 photographs here are light-microscope images of beautifully stained serial sections taken from one or another of a dozen species of flea. They are meant to illustrate the internal organs and tissues of insects in general, in a thorough mixture of morphology, histology and ultrastructure. The volume is aimed at students of disease-carrying insects or those beginning research on a wide range of insect biology, "carefully selected potted information which we would have appreciated enormously."

They remind us that it may well be that most animal species are parasitic forms, sufficient reason to flee an introduction to insect structure. The general reader will enjoy a time of thoughtful inspection, although the presentation in detail here of flea tissues organ by organ, from antennae and salivary glands to eyes and egg-surface apertures, seems specialized enough, even if "frankly elementary." There are good captions aided by auxiliary drawings and careful labels on the photographs.

On one page we see a flea jumping in close-up. The images are outlines drawn from five cinema frames taken at intervals of about a third of a millisecond; they begin at the start of the coordinated motion. The flea does not initiate its 150-*g* takeoff from its slender feet; it employs a sturdier bearing surface on the thigh. The sudden release of elastic energy stored in a special pad of resilin, a rubbery protein, accounts for the abruptness of the jump; that structure is examined in a number of microviews.

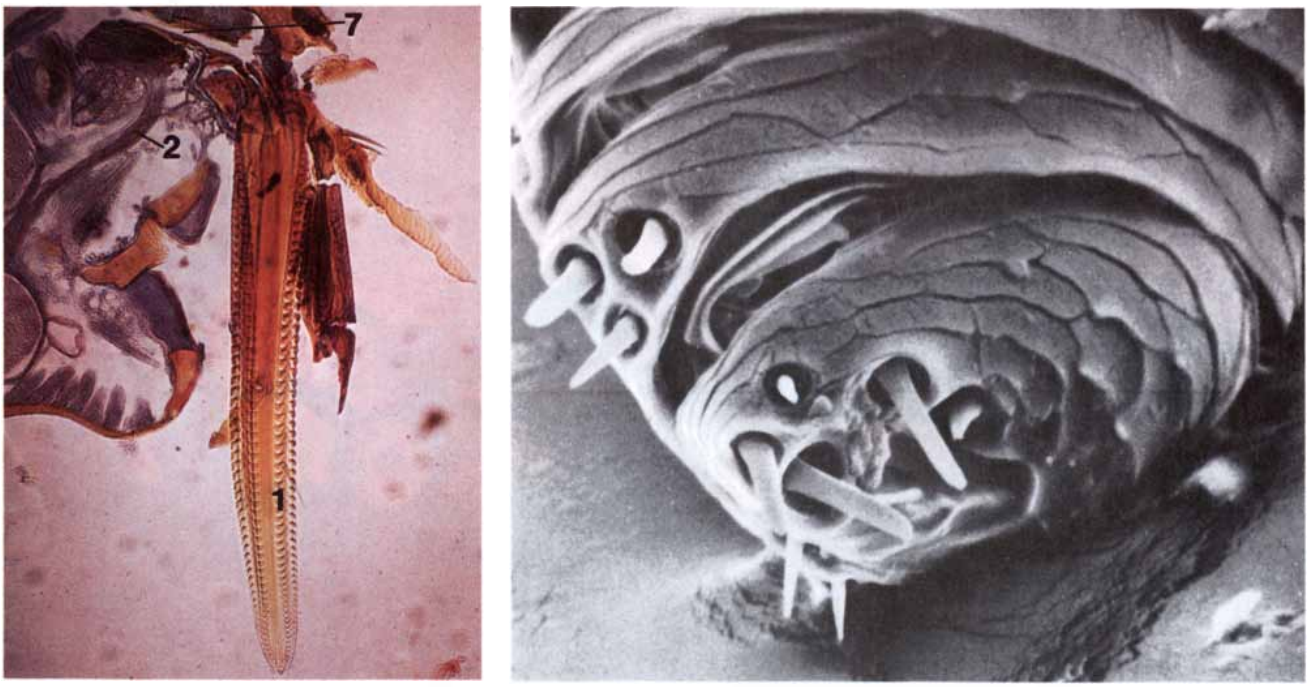
There are both transmission and scanning electron micrographs that show features at different scales, for example flea spermatozoa and eggs. Half a dozen photographs of other insect forms present important structures not found in fleas, the large compound insect eye, the ordered ultratexture of powerful flight muscles, and the scales of a butterfly.

These parasitologists do not fail to exhibit fleas with smaller parasites "that on [them] prey." Here are tiny mites near the oviduct of a flea, "massive coils" of a parasitic worm nearly filling a flea gut, a heavy internal infection by flagellates, and another by intracellular bacterialike rickettsiae. Worm cysts seen in a flea brain slow down the creature's electronics, so that it becomes easier prey for the parasite's final vertebrate host.

The last photograph of all evokes empathy. In it we see the intestinal tract of a ground-squirrel flea fatally blocked by a mass of rods of the plague bacillus, that single disease suffered alike by fleas, rats and humans. Dean Swift's witty verse is not wholly fulfilled; not even one still smaller flea is shown here on these smaller pests to bite 'em, although some analogue certainly exists. The flea's tormenting worm might itself have harbored a mite with its own infection of flagellates whose cells were host to some bacterial virus; that would be a living fivefold Chinese box to admire, all ordinary organelles apart. (It is the atoms, of course, that quench forever Swift's logical *ad infinitum*.)

In the September issue of Scientific American *Lester C. Thurow, professor of economics and management at the Massachusetts Institute of Technology, reviewed* The Positive Sum Strategy (*National Academy Press*). *The editors of the book, Ralph Landau, a former industrialist and consulting professor of economics at Stanford University, and Nathan Rosenberg, professor of economics at Stanford, have written the following reply to Thurow's review.*

We regard it as an excellent tradition that authors should not respond to reviews of their books. It is nevertheless one with which we will break. We believe the 41 individual



***Mouthparts* (left) *and tip of antenna* (right) *of* Spilopsyllus cuniculi, *from* A Colour Atlas of Insect Tissues Via the Flea**

chapters comprised by *The Positive Sum Strategy* address some of the most significant issues that confront the American economy. Successful policy will, we believe, turn in a critical way on a correct analysis of the economy's present ills. Yet Professor Thurow's characterization of the views of the 43 authors who contributed to *The Positive Sum Strategy* could well obscure what these economists, businessmen, scientists and technologists have to say. This is decidedly odd because there are actually significant areas of agreement between us and Thurow. Indeed, these areas of agreement are so substantial that we have some difficulty understanding the intensity with which Thurow states his disagreement with what we regard as the central themes of the book.

For example, he writes: "When it comes to one central theme of the book, the authors are certainly accurate: technology can be harnessed to raise living standards, while minimizing the use of labor and raw materials, so that we can escape from a zero-sum society in which every person's income gain must be matched by someone else's income loss." There is also a broad consensus among the authors that, for whatever reasons, our rates of saving and capital formation have been unacceptably low. On this critical issue Thurow must also be in agreement, because he states: "Nations or industries that invest more usually adopt newer technologies faster. The U.S. is simply investing less in acquiring new tools than its competitors do. It seems obvious that the U.S. will have to invest more." We are delighted that Thurow supports us on these two fundamental issues, since they are the two most central propositions of *The Positive Sum Strategy*. On what basis, then, can he still assert that "although many of the authors offer genuine insight, collectively they have, I think, got the diagnosis more wrong than right"?

The central thrust of Thurow's criticism begins with the attribution of an exceedingly high degree of uniformity to the individual authors and editors. He attributes to them the view that our economic problems are overwhelmingly due to an excessively intrusive government. He sees the book as expressing the view that our recent problems, especially those associated with a slow growth in productivity, are a result of a failure to "let the market work its magic." This is an exceedingly poor caricature of the contents of the book, a central feature of which is the expression of a wide diversity of views from people of very different backgrounds in business, science, engineering and academic life. One may criticize many of the present policies of the Government without necessarily being committed to the view that markets are perfect in their workings. There is a disturbing lack of subtlety in his argument, an attempt to force everybody onto some procrustean bed from which they presumably must swear allegiance either to complete laissez faire or to belief in a purely benevolent and totally efficacious government. Surely one may take issue with a number of government actions without necessarily believing the summation of economic wisdom is simply to "let market forces rip."

Thus Thurow's entire approach to policymaking is based on a highly polarized view of possible courses of action. This is a shame because an economist, of all people, ought to be asking the question of what would be the effects of certain adjustments at the margin. How would the performance of the economy change as a result of marginally raising or lowering certain magnitudes? An economist ought also to be sympathetic to the attempt of individual authors to communicate a sense of the wide diversity of circumstances that characterize different sectors of the economy and may call for very different kinds of government policies. Instead Thurow seems to be approaching the book by asking simplistically and divisively "Which side are you on" rather than telling readers what the book is about.

One example of the way in which this ideological approach suffuses the review must suffice. In his determination to portray the role of government intervention as benign, he offers rather extravagant praise for "administrative guidance" as the key to Japan's economic success story. Thurow is entitled to hold such a view, but he ought at least to inform his readers that the relevant chapter by Daniel I. Okimoto of Stanford tells an entirely different story. In evaluating the role of the Japanese government in influencing the performance of the high-technology industries, Professor Okimoto had stated: "The government's power to foster innovation is not nearly as great as is sometimes assumed.... It can identify the high-priority technologies, to be sure, but it cannot guarantee that private corporations will succeed in commercializing them. It can throw up a cordon of infant-industry protection around domestic markets in order to keep foreign competitors out, but that often has the perverse effect of dulling incentives to innovate. It can channel subsidies into 'targeted' R&D, but that can lead to distortions and waste in aggregate R&D investments. What lies within the government's effective power is largely limited to the creation of a healthy environment for business growth...." This is not exactly a testimonial for a policy of dirigisme and "administrative guidance," which Thurow appears to favor.

Apparently Thurow is anxious to absolve the Government of any substantial role in accounting for the deteriorating performance of the American economy in order to clear the decks for his own particular villain, American management. In his own words: "The fault does not lie with American government; it lies precisely with American firms.... Whatever else they may be, American managers are not world class when it comes to managing themselves."

It is no part of our intention to defend American management against Thurow. In fact, we find it easy to agree with a number of his criticisms of American managerial decision making, as do several of the authors of individual chapters, and we certainly heartily applaud his suggestion that Americans have much to learn from the Japanese in such matters. Indeed, the last three chapters of the volume deal entirely with the Japanese experience, but Thurow, curiously, ignores the contents of those chapters almost completely.

Rather, what we find both astonishing and perplexing is the total failure to link up the specific content of American managerial decisions with any aspects of the American economic environment—especially the macroeconomic environment that is inevitably shaped in significant ways by government. Thurow writes about these matters as if American managers are engaged in some kind of collective suicide pact, or are perhaps behaving like lemmings scampering into the sea. One need not be an economic determinist to believe there are powerful economic forces shaping the strategic decisions of American management. Thurow offers no intimation of what these might be.

As a major exhibit for his thesis Thurow offers the rapid growth in the number of managers and the concomitant slide in office productivity. He professes to believe "the average American office...is where the major problems lie." But what exactly has driven the drastic increases in white-collar employment in the past decade or so? This is a question of profound importance that is central to the ongoing structural changes in the American economy. In fact, two of the chapters in *The Positive Sum Strategy* (those by Stephen Roach of Morgan Stanley and James Brian Quinn of Dartmouth)

deal precisely with some of the complexities of this issue. Yet Thurow appears to be so involved in polemical exercise that he ignores these discussions and can see nothing beyond sheer managerial incompetence. Surely the answer must lie deeper than that. Is there really no competitive process at work that would sort out companies that are making poor decisions—for example, adding unnecessary layers of white-collar personnel—from companies that are making good ones? Is the lemming hypothesis of American management really so persuasive that one need not even consider alternative explanations?

The fact is, of course, that certain changes in the composition of the American labor force that are so thoroughly deplored by Thurow, and cited as presumptive evidence of American managerial incompetence, are merely recent expressions of long-term trends going back at least to the beginning of the 20th century. The rise in white-collar employment as a percentage of the labor force is in no sense (as Thurow appears to suggest) a recent phenomenon. It is closely linked to complex changes in the economy—changes in the composition of demand, in technology, business organization, patterns of industrial specialization and the spatial distribution of companies and households, to cite some of the major factors.

Obviously Thurow is right in asserting that productivity levels would be higher if we could obtain the same levels of G.N.P. with smaller numbers of white-collar employment; such a tautology is not very illuminating. The serious questions are: What are the forces driving the growth in white-collar employment, and why is it apparently not more closely linked with greater growth in G.N.P.? To invoke managerial incompetence, with a rhetorical flourish, as the major explanatory variable is to trivialize some very deep and perplexing economic issues.

Similarly, Thurow's determination to pin the tail of poor productivity performance on the managerial donkey leads to a total neglect of one of the most drastic social changes in American history since the Second World War. This was, of course, the demographic revolution that included the maturing of the postwar baby-boom cohort and the steady surge of participation in the labor force by women, as well as a sizable flow of immigrants. In the 30-year period from 1955 to 1985 the American economy managed to adapt to the influx by creating no fewer than 45 million new jobs. In comparison, the Common Market countries created only five million during the same period.

This was a remarkable accomplishment, but one for which a price was paid: the rapid expansion in total numbers and the qualitative deterioration associated with a flood of inexperienced workers into the labor force inevitably had a serious negative impact on productivity. Furthermore, a large fraction of these new entrants moved into the service sectors, especially those specific sectors where employment could readily be increased with less capital and less stringent demands for human skill than in other parts of the economy. It would seem difficult to discuss poor productivity performance and the rapid growth of employment in services over the past 30 years without even a mention of these fundamental economic conditions, but Thurow has managed to do so. He complains about "a work force low in quality" without a single reference to these underlying changes.

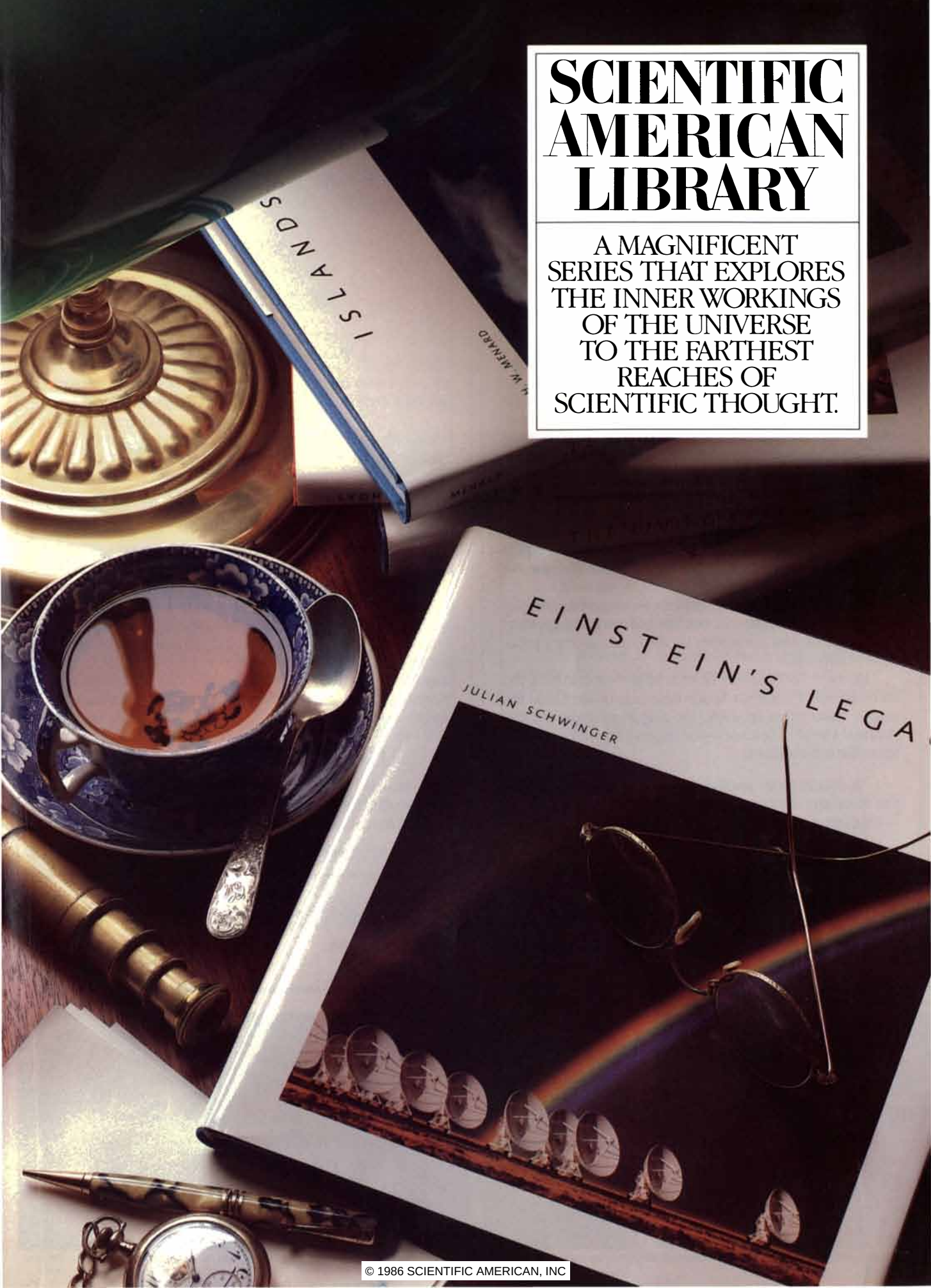
Few economists, we also believe, would wish to deny that government policies with respect to spending and taxation have had a great deal to do with the high interest rates and low savings rates that have prevailed in America in recent years. The high interest rates have been a pervasive factor in discouraging business investment and in imposing short time horizons on business managers. One does not need to consult a compound-interest table to understand how double-digit interest rates particularly discourage long-term investment and lead to the preoccupation with quarterly reports and insistence on short-term payoffs for which American management has so often been criticized. Japanese businessmen do indeed take a longer-term view, but there are excellent economic reasons for this, such as the much higher savings rate in Japan and associated differences in the cost of capital between the two countries. While we would not want to deny the possibility of benefits flowing from a greater degree of participatory management, surely large differences in the cost and availability of capital (a subject to which the last chapter of the book is devoted) deserves at least a single mention.

Thurow himself admits that insufficient investment by American business is indeed a serious economic problem. Surely such an admission points directly to the significance of government tax structures that systematically favor consumption over saving (in sharp contrast to the Japanese practice), and to government budgetary practices involving large Federal deficits (that is, dissaving) and consequent upward pressures on interest rates. It is difficult to see how a discussion of the recent state of the American economy, with its huge trade deficits and an overvalued dollar, could possibly ignore these issues or cavalierly dismiss their significance.

It is also difficult to understand how Thurow can compare America's poor recent economic performance with the more favorable performance in the years after World War II without at least acknowledging the rather special circumstances that prevailed at that time—specifically the devastation inflicted on our commercial competitors in Europe and Japan by the war. Managerial decisions never take place in a vacuum, and a major responsibility of economics is—or ought to be—to show how such decisions are shaped by the macroeconomic environment. But such a discussion might suggest that an important source of our recent difficulties lies at the door of government policy and is not attributable to incompetent management.

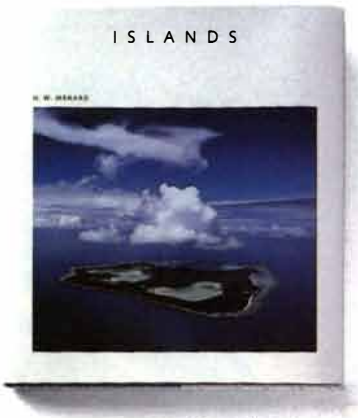
We believe our book offers a genuinely serious discussion of the many complex issues involved in harnessing technology for economic growth, and we are confident that the reader will gain fresh and useful insights from it. One of our major points was that gaps in the knowledge of the microeconomics of technological change require much more intensive research at a disaggregated level. It was for precisely this reason that we invited a heterogeneous collection of economists, technologists, scientists and business executives. We were anxious to present a wide range of views that would also reflect the divergent needs and situations of various sectors of the economy.

In spite of this diversity we did discover a broad consensus that an improvement in the long-term growth rate required an increased rate of spending on R&D, greater efforts to improve the quality of the labor force through education and training, and a more rapid rate of growth of the capital:labor ratio. Perhaps the most hopeful findings came from distinguished scientists such as William O. Baker, the retired chairman of the Bell Laboratories, and James D. Watson, director of the Cold Spring Harbor Laboratory, who suggested that major new technologies were already looming on the horizon and that therefore there are likely to be great opportunities for accelerating productivity and growth in the future. Capitalizing on these opportunities, however, will require substantial policy changes in *both* the public and the private sectors.



ISLANDS

H.W. Menard



In a lifetime of scientific adventure on the broad reaches of the Pacific Ocean, H.W. Menard discovered in the Earth's most romantic islands new insights into the forces deep in the Earth's interior.

His central contributions to plate tectonics, the unifying theory of geology, explain the layout of the island arcs, the atolls and the seamounts on the map of our globe. He also tells the story of the evolution of life, of human settlement and of the exploration, in the latest centuries, of these islands.

Sharing in Menard's adventure, the reader will also share his understanding of the Earth as a dynamic system. Photographs and cross section diagrams illuminate the text and make vivid connection between the beauty of nature at the surface and the processes driving the system in its depths.

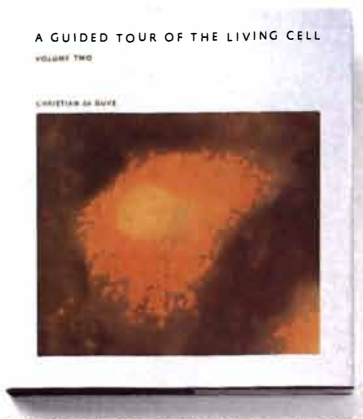
H.W. Menard was a founding member of the Scripps Institution of Oceanography, where he was professor of Geology. In the course of his distinguished career, he served as Technical Advisor in the President's Office of Science and Technology and as Director of the U.S. Geological Survey.

"In a very real sense, this book is his scientific legacy."—ROGER REVELLE, PROFESSOR OF SCIENCE AND PUBLIC POLICY, UNIVERSITY OF CALIFORNIA, SAN DIEGO

230 PAGES, 152 MAPS, PHOTOGRAPHS AND ILLUSTRATIONS

THE LIVING CELL

Christian de Duve



In A GUIDED TOUR OF THE LIVING CELL, Christian de Duve invites us to accompany him on an engrossing expedition through the trillions of cells that make up the human body.

The journey is divided into three itineraries. In the first, we examine a cell's outer and inner membranes and their two-way commerce with the exterior. In the second, we call at all the major organelles—the cell's own organs—and observe them manufacturing their own special products. Our third itinerary takes us into the nucleus itself, where we see DNA steering the reproductive biochemical machinery of the cell.

Christian de Duve is Andrew W. Mellon Professor at The Rockefeller University in New York. He was a joint recipient of the Nobel Prize for medicine in 1974.

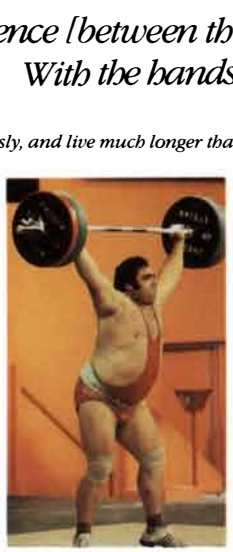
"It is hard to imagine a better introduction to the world of the cell and all it means to everyone."

—JOHN BARKHAM REVIEWS

2 VOLUMES, 422 PAGES, 425 ILLUSTRATIONS

"The difference [between this and other book series] comes in the quality of
With the handsome volumes in front of us, what response can

The largest forms of life reproduce slowly, more ponderously, and live much longer than smaller organisms. On Size and Life *examines why.*



A molecular view of the plasma membrane. Electr through an axopod of the heliozoon Echinosphae

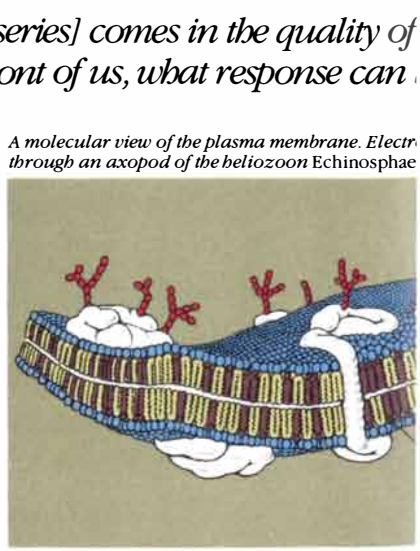
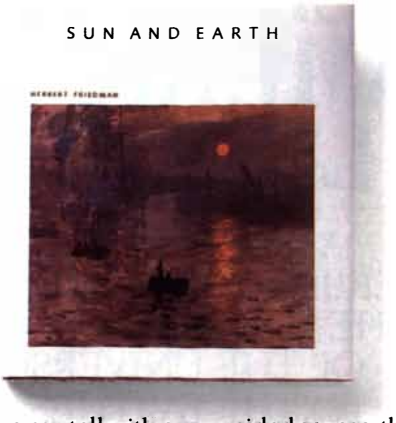


PHOTO CREDITS—Daphnia: James Bell, Photo Researchers, Inc./Ant: Daniel Adams/Elephant: Peter Johnson, National History Photographic Agency/Illustration drawn by Neil O. Hardy/Electron micro Mutual Research Corporation

SUN AND EARTH

Herbert Friedman



As we can tell with our unaided senses, the distant sun bathes the Earth in radiant energy. From instruments on the ground and in orbit, we have now learned that the Earth revolves around in the Sun's extended atmosphere—the two bodies are in contact. An out-rushing solar wind engages our outer atmosphere and sets into action the processes that reach all the way to the planet's surface.

In SUN AND EARTH, the space-scientist Herbert Friedman combines an historical narrative with interpretive accounts that retain the awe and wonder with which scientists and poets have regarded the sun's life-bestowing characteristics.

SUN AND EARTH also provides a comprehensive picture of how the sun shapes our environment and explains the impact of solar variability on the contemporary problems of survival.

Herbert Friedman is chief scientist emeritus of the E.O. Hulburt Center for Space Research of the Naval Research Laboratory. In 1969, Dr. Friedman received the National Medal of Science—the nation's highest honor for scientific achievement.

"It covers more ground in an authoritative fashion than any contemporary book I've seen."
—WILLIAM C. LIVINGSTON, ASTRONOMER

251 PAGES, 195 ILLUSTRATIONS

OTHER VOLUMES IN THE LIBRARY:

SUBATOMIC PARTICLES
Steven Weinberg

MATHEMATICS AND OPTIMAL FORM
Stefan Hildebrandt and Anthony J. Tromba

POWERS OF TEN
Philip and Phylis Morrison and the Office of Charles and Ray Eames

HUMAN DIVERSITY
Richard Lewontin

PERCEPTION
Irvin Rock

DRUGS AND THE BRAIN
Solomon H. Snyder, M.D.

THE SECOND LAW
P.W. Atkins

THE SOLAR SYSTEM
Roman Smoluchowski

THE SCIENCE OF MUSICAL SOUND
John R. Pierce

FOSSILS
George Gaylord Simpson

CONSTRUCTING THE UNIVERSE
David Layzer

FIRE
John W. Lyons

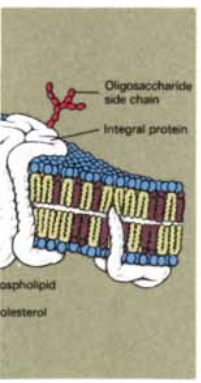
ON SIZE AND LIFE
Thomas A. McMahon and John Tyler Bonner

THE TIMING OF BIOLOGICAL CLOCKS
Arthur T. Winfree (in preparation)

EXTINCTION
Steven M. Stanley (in preparation)

authors, who combine eminence in their fields with proven gifts of exposition.
e be but immediate surrender to this firepower?" —Nigel Calder, NATURE

rograph of a section
ucleofilum. From The Living Cell.



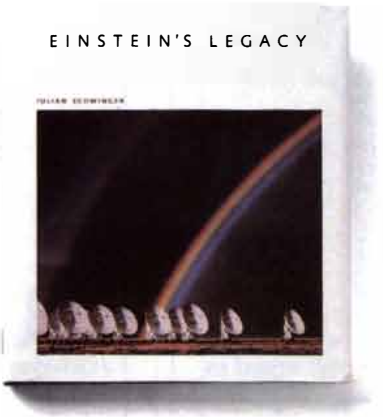
The "crowning" of a forest fire. Laser-beam spectroscopy. The copper converter. An experimental warehouse fire. From Fire.

ourtesy of Dr. L.E. Roth and Dr. Y. Shigenaka/Fire in trees: L. Burr, Gama-Liaison/Copper Smelter: John Moss, Photo Researchers, Inc./Laser: Courtesy of Sandia Labs/Plastics Fire Test: Courtesy of Factory

Your Premier Volume

EINSTEIN'S LEGACY

Julian Schwinger



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. Technical demands on the reader do not exceed the most elementary algebra.

"It is particularly gratifying that the subject's intricacies, including glimpses of non-Euclidean and Riemannian geometries, and the theory's early successes, . . . 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

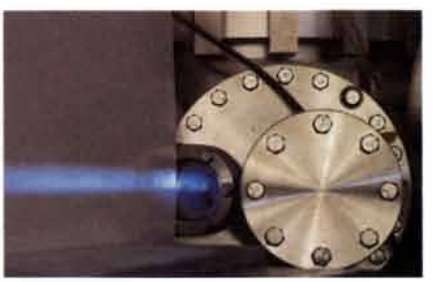
250 PAGES, 200 ILLUSTRATIONS

COVER PHOTO: The rainbow arches over the Very Large Array radio telescope of Soccoro, New Mexico. For nearly three generations, Einstein's legacy has withstood experimental test by instrumentation not yet invented when Einstein did his thinking. © Doug Johnson

TO GUARANTEE SWIFT DELIVERY OF YOUR FIRST VOLUME OF THE SCIENTIFIC AMERICAN LIBRARY, WE WILL SHIP *EINSTEIN'S LEGACY* VIA UPS.

Radio telescopes
Stephen Northup/Black Star © 1980

Einstein and Chaplin in Hollywood, 1931
The Granger Collection



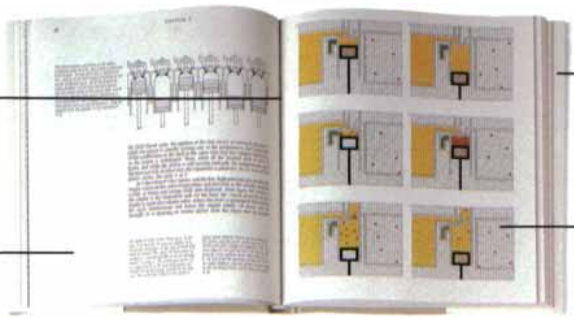
Synchrotron light produced at Brookhaven National Laboratory

Now in its third year, SCIENTIFIC AMERICAN LIBRARY has won recognition as an exciting new genre in the literature of science. Created out of the collaboration of scientists and editors that distinguishes SCIENTIFIC AMERICAN itself, each volume in the LIBRARY:

☐ Covers a topic of major interest with immediate relevance to ongoing research in its own and related fields.
☐ Carries the authority of an author who has made a significant personal contribution to its subject.
☐ Proves there is no topic in any field of science that cannot be explained to readers who want to understand.

Impeccably printed on high quality paper. Heavy cloth covers and binding to insure years of trouble-free reference.

Ample margins for visual relief and notes.



Large, inviting format (8½" x 9¼") opens flat for convenient reading.

Handsomely illustrated with many full-color charts, diagrams, maps and photographs.

JOIN OUR LIBRARY. THERE IS NO RISK INVOLVED.

A Two-Month Cycle.
As a library member you will receive a new volume approximately every two months. (Your first volume, EINSTEIN'S LEGACY, will be shipped UPS for fastest delivery.)

15-Day Free Examination.
Examine each new volume at your leisure. If you decide within 15 days not to own a particular volume, simply return it, and your account will be promptly credited.

No Prepayment Necessary.
Pay only for the volumes you decide to keep.

No Minimum Purchase.
You may choose to keep as many or as few volumes as you like. There is no yearly minimum.

Low Members' Price.
Members may purchase any volume at the exclusive low price of $24.95—well below the bookstore price.

Your Guarantee.
You may cancel your membership at any time without notice, for any reason, without penalty.

MEMBERSHIP RESERVATION

Free with your paid order: TELESCOPING POCKET TRAVEL ALARM.

Its sleek black case slides open to reveal a 7-function alarm and calendar clock with a dependable quartz movement. Comes with mini-support leg for your viewing convenience.

Please speed my premier volume—EINSTEIN'S LEGACY—on its way via UPS and enroll me as a member of the Scientific American Library. Thereafter, send me a volume approximately every two months for 15 days' free examination.

I need send no payment now. For each volume I keep, I will be billed $24.95 plus postage and handling. This price is guaranteed for at least one full year. I understand that I may cancel my membership at any time, for any reason, and without penalty.

Yes, I'm looking forward to receiving my LCD Travel alarm after you receive payment for my first volume.

Signature _____

Name _____

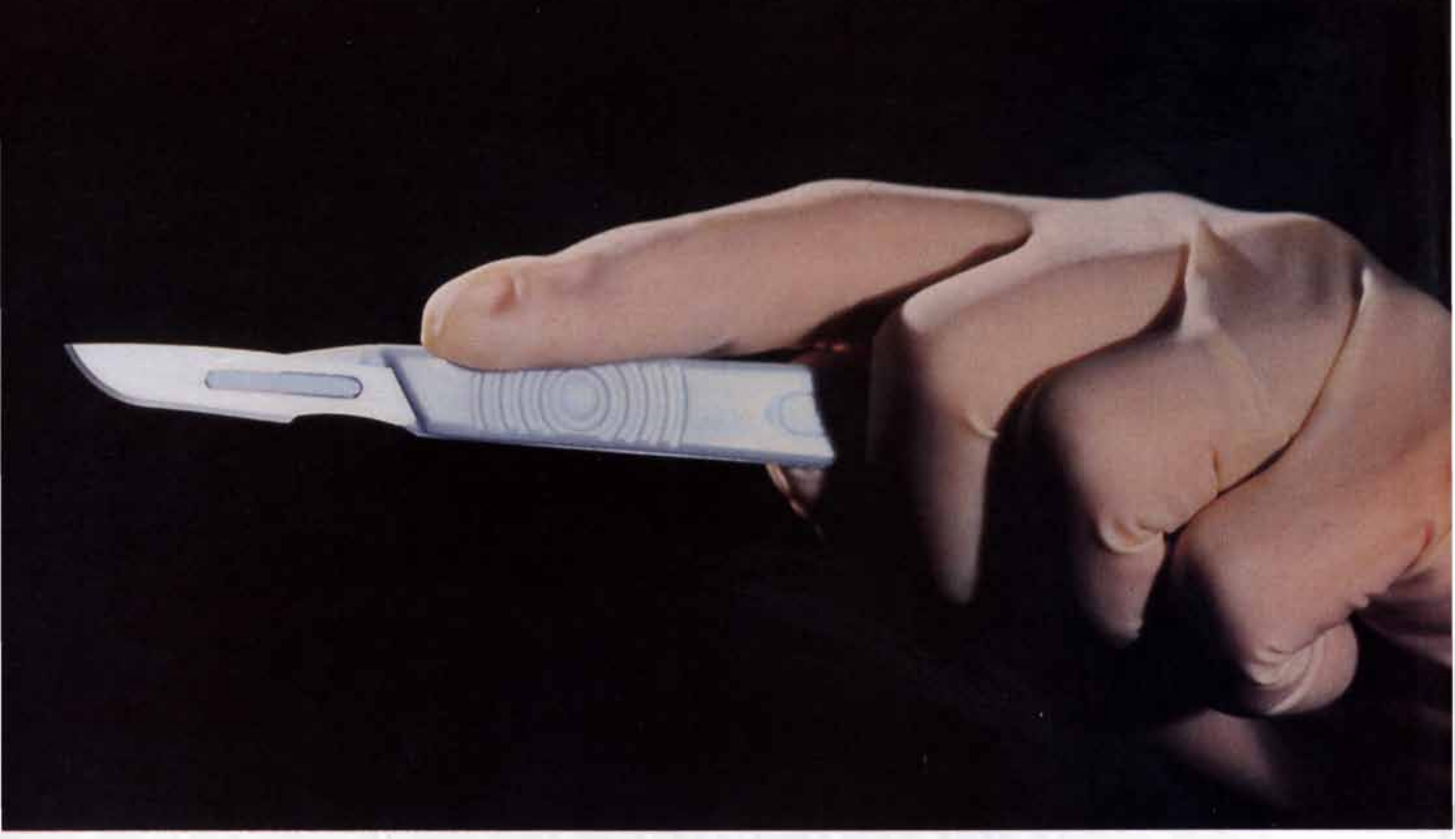
Address _____

City _____ State _____ Zip _____

3PAP

SCIENTIFIC AMERICAN LIBRARY

Mail this form to: SCIENTIFIC AMERICAN LIBRARY, P.O. BOX 646 HOLMES, PA 19043.
Or, to join right away call toll-free 1-800-345-8112 (1-800-662-2444 in Pennsylvania).



Until recently, this was the leading edge.

The leading edge of technology is replacing the surgeon's scalpel in the removal of kidney stones. Today, almost 90 percent of kidney stone sufferers have an alternative—the Lithotripter.

Developed by Dornier, the Lithotripter disintegrates kidney stones through shock waves. Described as a "medical miracle" by the Secretary of Health and Human Services, the Lithotripter eliminates the pain and discomfort normally associated with surgical procedures. And because it is completely noninvasive, the Lithotripter significantly shortens recovery time, thus reducing cost to the patient. Thus far, over 100,000 kidney stone sufferers have been successfully treated by the Lithotripter.

Dornier developed the Lithotripter as a technological refinement of research which had originally been targeted toward studying the effect of shock waves on supersonic aircraft. Lithotripter—it's the only device of its kind in the world, working today for a healthier future.

DORNIER

DORNIER MEDICAL SYSTEMS, INC., 824 LIVINGSTON COURT, MARIETTA, GA 30067

Established 1845 SCIENTIFIC AMERICAN November 1986 Volume 255 Number 5

The Shadow Market in Housing

It consists of the processes that provide housing by means other than new construction. If state and local laws gave this activity more scope, many affordable dwelling units would be available

by William C. Baer

The usual solution proposed to deal with the chronic shortage of affordable housing in the U.S. is to build new housing, often with Federal subsidies to reduce the cost to the buyer or renter. Hardly anyone thinks of encouraging more activity in what has been called the shadow market. It consists of the many processes that reconstruct the existing stock of housing by subdividing, merging, enlarging and generally improving it and that increase the stock by transforming nonresidential structures into housing. The many forces that lead to these effects are local, and it is to local governments that one should look for moves to expand the already significant impact of the shadow market.

Some indication of that impact appears in the figures for 1980 as assembled by the U.S. Department of Commerce. In that year, while developers spent some $47 billion on new housing, owners of residential property spent another $46 billion on existing housing for maintenance, additions and alterations. Data on new housing are well publicized as an important economic indicator, but little mention is made of the information on maintenance and remodeling; virtually nothing is said about additions to the housing stock from sources other than new construction. Yet these activities accounted for 21 percent of the increase in the total housing stock from October, 1973, through September, 1980.

The shadow market was even more important as a source of low-cost housing in particular. From 1973 to 1980 it accounted for one third of the low-cost units occupied by owners and half of the units occupied by renters. During the same period 85 percent of the low-cost rental housing that was built new rather than reconstructed was subsidized. Subsidized housing is extremely vulnerable to budget cuts in the face of competing demands for funds and the prospect of a continuing Federal deficit. Subsidized new construction would seem to be an unreliable source of low-cost housing for the remainder of the 1980's and perhaps throughout the 1990's. On the other hand, most of the low-cost rental housing supplied by the shadow market was unsubsidized (again some 85 percent), making that market a dependable source and a logical focus for housing policy.

The importance of providing low-cost housing is demonstrated by the shift in the nation's housing problem over the past four decades. In 1940 the housing stock was beset by problems of quality; today the problem is that much housing strains the resources of many families or is quite beyond their means. Physical inadequacies in the stock have been significantly alleviated since 1940. For example, 40 percent of the dwellings in 1940 lacked some or all plumbing facilities, whereas in 1980 the figure was less than 5 percent. Overcrowding, defined as more than one occupant per room, declined from about 20 percent to about 4 percent in the same period.

Meanwhile the cost of housing rose steeply with respect to the traditional standard that a renter should devote a fourth or less of his or her income to this item. In 1950 some 32 percent of renters paid more than a fourth of their income, but by 1980 more than 50 percent were doing so. Increasing the shadow market's contribution to low-cost housing is an obvious response to the problem of affordability.

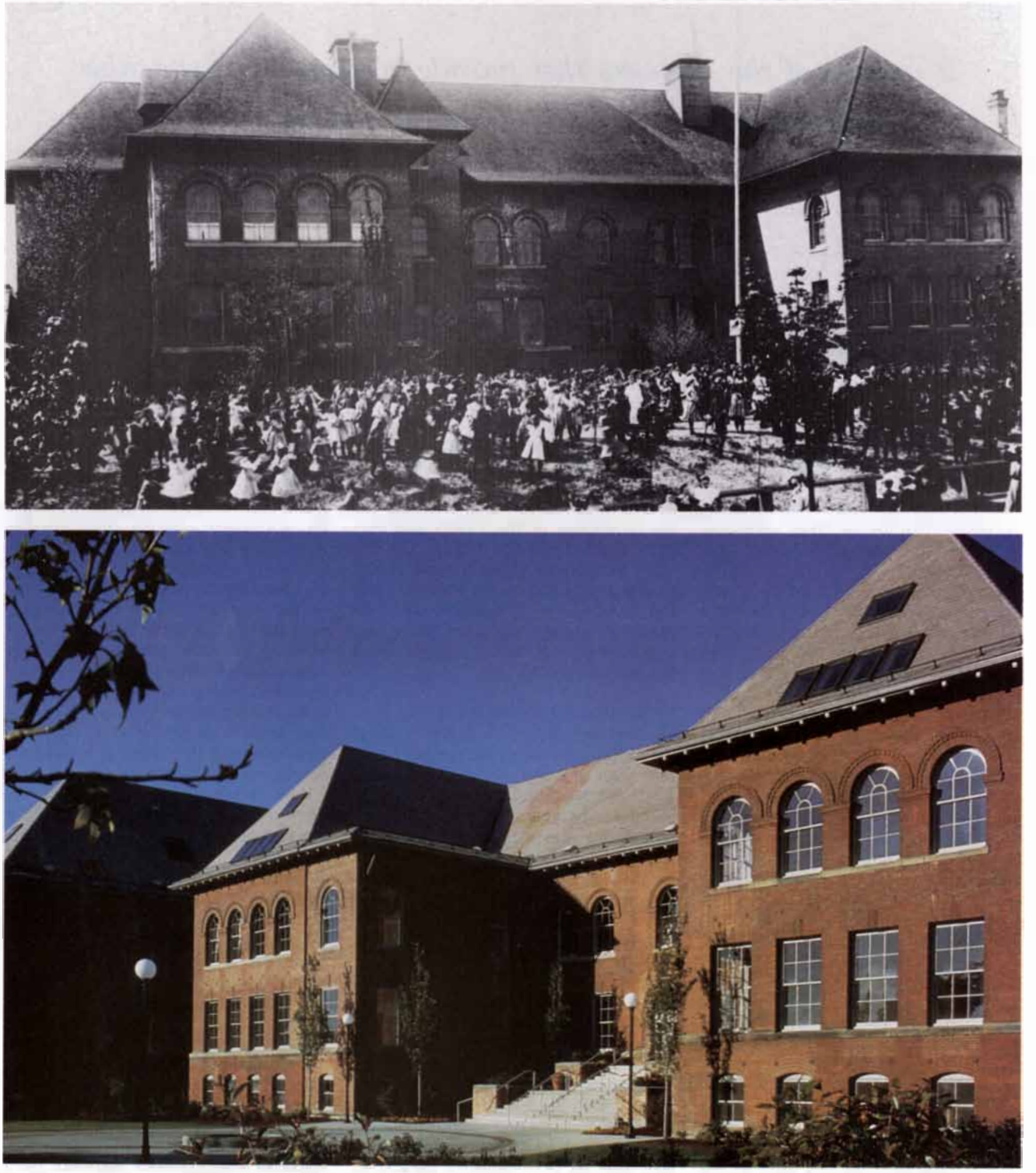
Because it is a local phenomenon, the shadow market is not dependent on national housing policy and funding priorities. Instead it is most susceptible to zoning and land-use regulations—a local-government function. By selectively altering these regulations local governments could significantly increase the effect of the shadow market. Existing housing would be more readily adapted to changing patterns in demand. More low-cost housing would be provided with only minor public expenditures. To demonstrate these possibilities requires an examination in greater detail of the comparative importance of new construction and the shadow market as sources of low-cost housing.

The conventional view that additions to the U.S. housing stock occur through new construction needs modification. New construction is indeed the main source of additions, but it is becoming less dominant. During the 1950's and 1960's about 90 percent of all additions came from new construction, but in the 1970's that source dropped below 80 percent, reaching an average of 73 percent in the years from 1973 to 1980. The shadow market accounted for the remainder.

Behind these averages is a considerable annual fluctuation. The production of new housing can vary greatly over a decade, depending on interest rates, the state of the economy and national demographics. Between 1970 and 1980 the number of housing starts ranged from a high of 2,379,000 in

1972 to a low of 1,171,000 in 1975. Similarly, and for the same reasons, the preponderance in the type of unit built varies, shifting for example between single- and multiple-family dwellings.

Notwithstanding these conventional statistics, which monitor annual and cumulative change, the 1980 census of housing revealed one million more units than could be accounted for by new construction. Most of the increase was in rental units and was accounted for by the shadow market. I should not leave the impression that these unaccounted-for units came as a total surprise. Since 1950 the U.S. Bureau of the Census has reported at 10-year intervals (in its summary of components of inventory change, usually abbreviated as CINCH) the contributions of non-new construction to the housing stock. Since 1973 the bureau has also collected some of this information periodically through its American Housing Survey. Although these two sets of data provide a glimpse of the dynam-



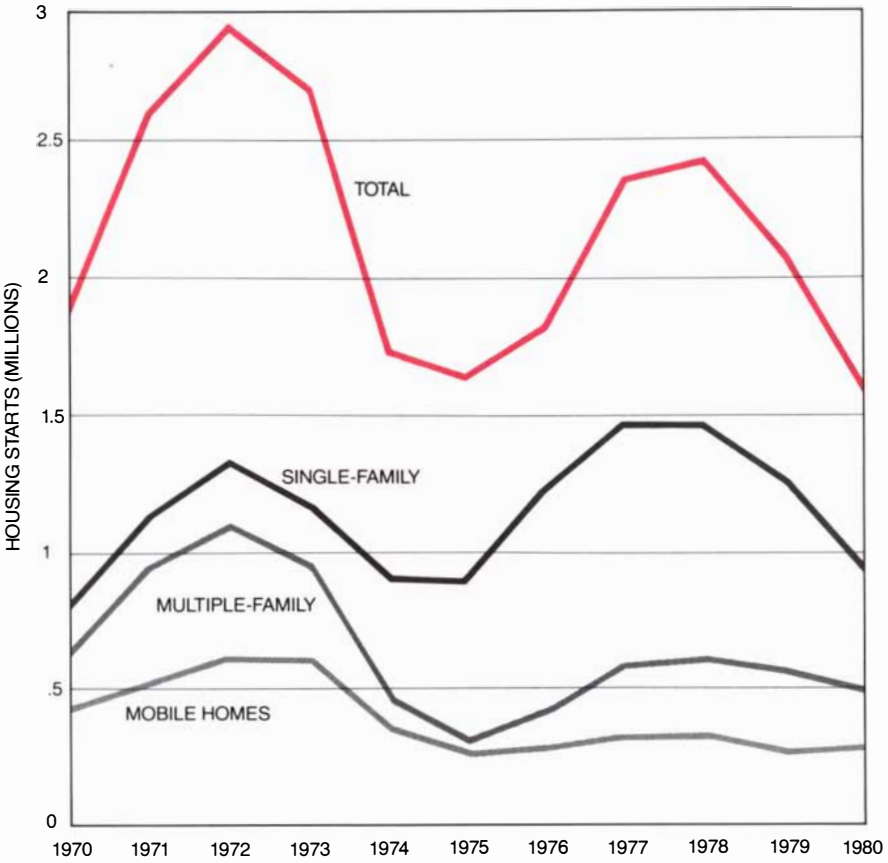
TYPICAL PROCESS of the shadow market was the conversion of a school in Seattle into an apartment building. At the top is the building when it was the Queen Anne School; now (*bottom*) it is a 49-unit condominium. The units cost from $52,000 to $195,000.

ics of housing, they are not widely publicized.

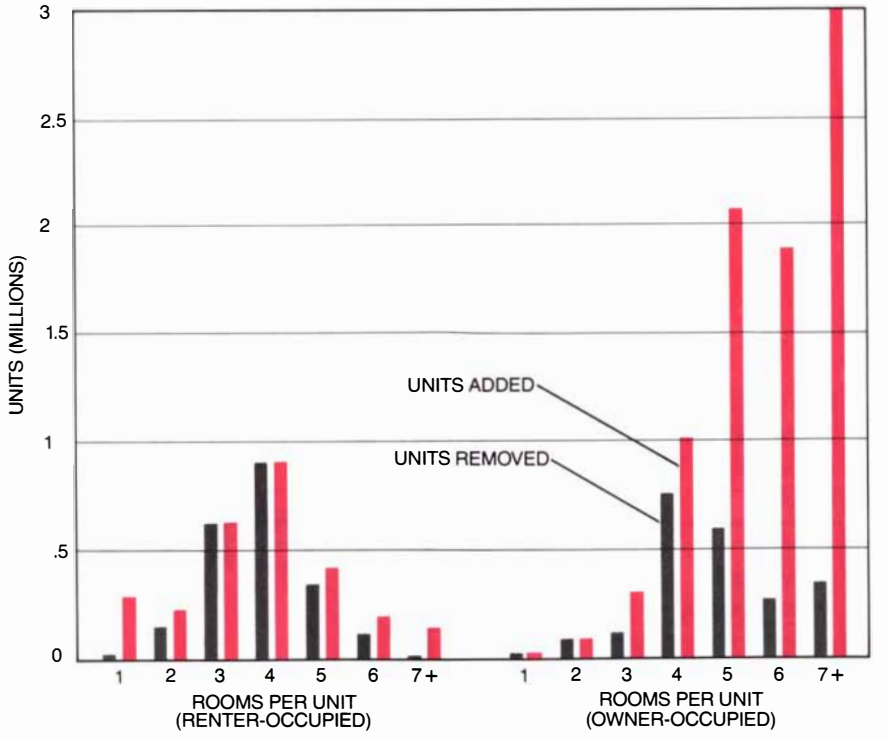
The dynamics arise in considerable part from the fact that housing is an extremely heterogeneous product, reflecting the uniqueness of each building site, the long life of the structure and particularly the decentralized means of production. Decentralization shows up sharply in a comparison with the automobile, which for most people is second only to housing in size and cost. Only a few manufacturers produce automobiles. In housing, however, the 100 largest developers produce only about 15 percent of the annual new construction. The remainder is mostly built by companies of small or medium size or by individuals—all with their own idea of what constitutes a house. Moreover, the builder, unlike the automobile manufacturer, makes no systematic effort to control the quality of maintenance once the product has been sold. Instead the owner must deal with an even more decentralized system of local repair and remodeling services—all with their own views of what should be done. Local building codes are one of the few sources of standardized construction and maintenance procedures, yet they are notorious for differing from one another and for being circumvented by owners in any event.

Decentralized maintenance makes it difficult to keep track of changes in the quality of housing. Concealed or visible, done illegally or officially sanctioned by building permits, the actual physical changes to the existing stock are not noted in a national data base when they take place. Physical inspection to enumerate the characteristics of housing is rarely done because it is expensive and time-consuming. Residents and owners can be asked to enumerate the characteristics of their housing, but the information is unreliable if objective assessments of structural condition are required.

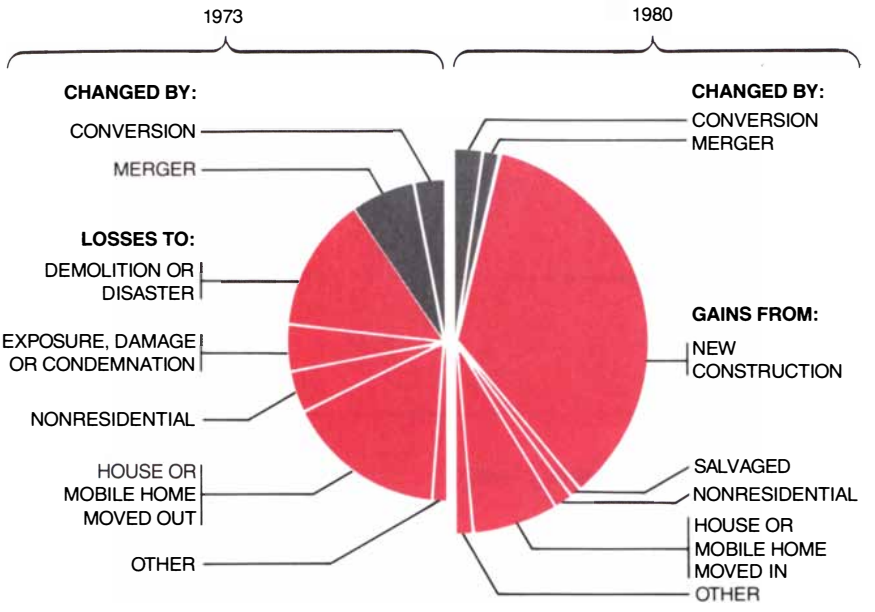
Hence the snapshots from the decennial census of housing provide most of the information about changes of quality in housing. Little is known of how the changes come about. One can, however, obtain a rudimentary understanding of housing dynamics from the CINCH data for 1973–83. The two most basic changes reported in the CINCH surveys result from additions to and removals from the housing stock. The CINCH data show that from 1973 to 1980 the removals amounted to about a third of the additions. Some 8.4 percent of the 1973 stock was lost during the period, while 20.2 percent of the 1980 stock (a larger base) came into being.



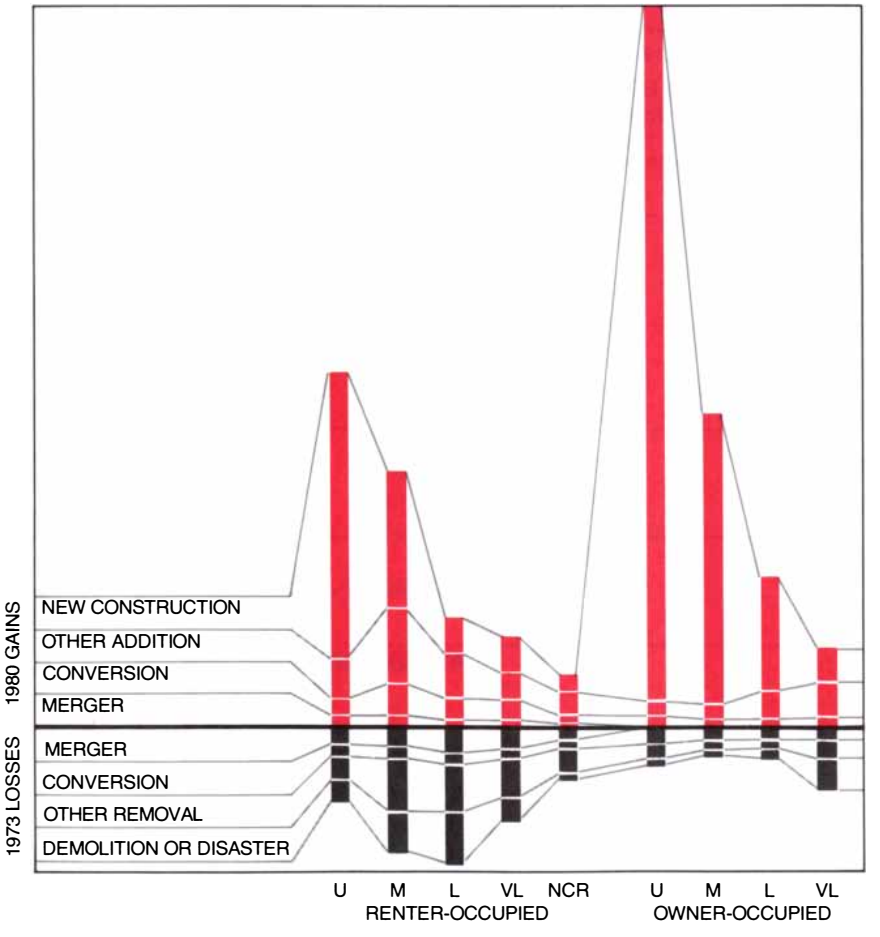
PRODUCTION OF HOUSING in the U.S. fluctuated considerably between 1970 and 1980. Each basic category reacted similarly to national economic influences. During the same 10-year period expenditures by owners of residential property for maintenance, repair and improvement rose steadily, showing some of the activity of the shadow market.



ADDITIONS AND REMOVALS in the housing stock from October, 1973, through September 1980, are charted. More units were added than were removed, and the median number of rooms added per unit was one more than the number removed. These trends have contributed to an improvement in housing quality and an increase in housing costs. The increase in rooms and units also provides a rich source for shadow-market activity.



HOUSING CHANGES were brought about by the shadow market from 1973 to 1980. The 1973 stock was 77.2 million units. Most of them (91.6 percent) remained the same; the fate of the others is charted at the left. By 1980 the stock was 89.3 million units (79.2 percent unchanged from 1973). The sources of the other units are charted at the right.



CHANGES BY VALUE affected high- and low-value units unequally. The value categories, represented by letters, are upper, moderate, low, very low and (for rental units) no cash rent. Conversion is the subdivision of a single housing unit into two units or more; merger combines two units or more into a single unit. These and other shadow-market activities also affected renter- and owner-occupied units to significantly different extents.

Most of the removals, however, actually were an accounting notation rather than a true loss. Only about a fourth of the dwellings were really lost as a result of demolition or disaster. The rest were "lost" because they were transformed into something else. Examples include conversion, merger and a change from residential to nonresidential use, as when a house becomes a restaurant or an office.

An examination of such transformations provides more detail on the functioning of the shadow market. Additions by the shadow market stem basically from three sources: the restoration of a previously uninhabitable unit; the transformation of a nonresidential space into a dwelling, and the conversion of group quarters (such as a sorority house) into individual residence quarters.

Restorations added only about 1 percent to the stock between 1973 and 1980. Transformations accounted for 3.2 percent. Discovering these transformations is difficult because they appear from unexpected sources and in unconventional places. The sources include garages, stores, barns and schools. The transformations add to the high-cost stock of housing as well as to the low-cost one. For example, public and private redevelopment of industrial waterfront areas changes obsolete warehouses and factories into expensive rental or condominium apartments.

Group quarters tend to be overlooked as a source of housing because they shelter "nonhousehold" people. Group quarters include prisons, college dormitories, military barracks and institutional and religious housing. Nevertheless, they become housing units when they are converted into, say, apartments. About 1.5 percent of all additions came from this source.

There are still other sources of additions and causes of removals, but they differ in that they simultaneously create additions and removals, although not necessarily in equal numbers. The sources include conversion, merger and moving a house or a mobile home to a new site. Conversion, which entails subdividing one unit into two units or more, results in additions that outnumber the removals. Between 1973 and 1980 the average conversion resulted in 2.4 units for every unit converted; conversions accounted for 3.1 percent of all additions. Mergers, which combine two or more units into fewer units, were responsible for 2.6 percent of all additions but removed two units for every unit merged. Moving a housing unit is also both a removal and an addition. It accounted for

15.7 percent of the 1973–80 additions and 38.3 percent of removals (calculated on the smaller base of 1973).

Conversions, mergers and moves are all examples of aggregate change, amounting to about 21 percent of all additions during the period. A more pertinent question for housing policy is differential change, meaning processes that work differently on low- and high-cost units. Are additions uniform in amount and percentage for all levels of quality or do they tend to provide mostly high-cost housing? Do removals occur equally across all levels of quality or is low-cost housing more likely to be removed?

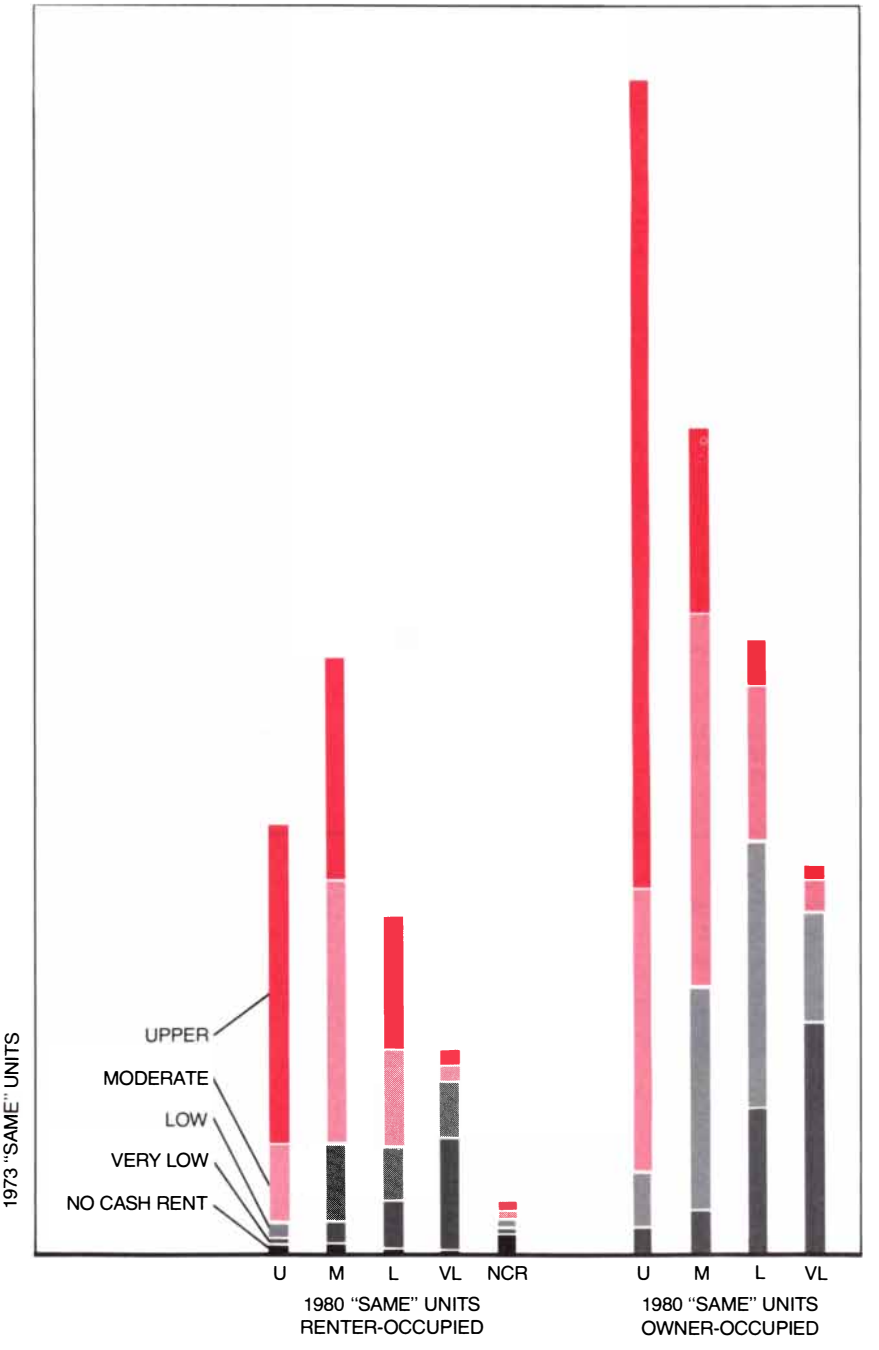
It is possible to construct from the CINCH data a model that allows answers to these questions. Given the problem of affordability that I have mentioned, the value of owner-occupied units and the gross rent for renter-occupied units provide a means of categorizing segments of the housing stock. Such information on cost is a good surrogate for the more complex characteristics of quality, and it is roughly congruent with income.

Invoking such classifications as upper and lower for incomes and housing always raises problems. There is no standard, universally accepted way of determining appropriate categories for these social concepts. The Federal Government and some state governments have, however, adopted a system of classifying household income to determine who is eligible for housing assistance. The same approach can serve to categorize four classes of housing value and rent: upper, which is greater than 120 percent of the median value or rent; moderate (between 120 and 80 percent of the median); low (between 80 and 50 percent), and very low (less than 50 percent). For renters a category of "No cash rent," which covers about 5 percent of all rental units, must be included. With these measures one can compare categories from one time period to another.

Three rather interesting findings emerge from this more detailed analysis. First, rental units were about five times more vulnerable to removal than owner-occupied units (11 v. 2 percent). The reasons for removal also differed between owner- and renter-occupied units. In the owner-occupied group the units in the upper category of value tended to be removed by transformation into a more intense use (mainly conversion into multiple units), whereas units in the low-value categories tended to be eliminated altogether. In the rented group mergers were a significant (18 percent) cause of removal, more so in the low categories than in the others. Apparently two rental units that might otherwise be demolished can often be combined into a larger unit and thereby remain competitive.

Second, as one might expect, the increased vulnerability to loss was skewed toward owner- and renter-occupied units in the lower categories. For example, the very low category for renters sustained a loss rate four times as high as the loss registered for the upper category.

Additions to the stock also exhibited major differences between owner- and renter-occupied units and between cost categories. One difference is that most additions to the housing stock start out as owner-occupied units. There were only about two-thirds as

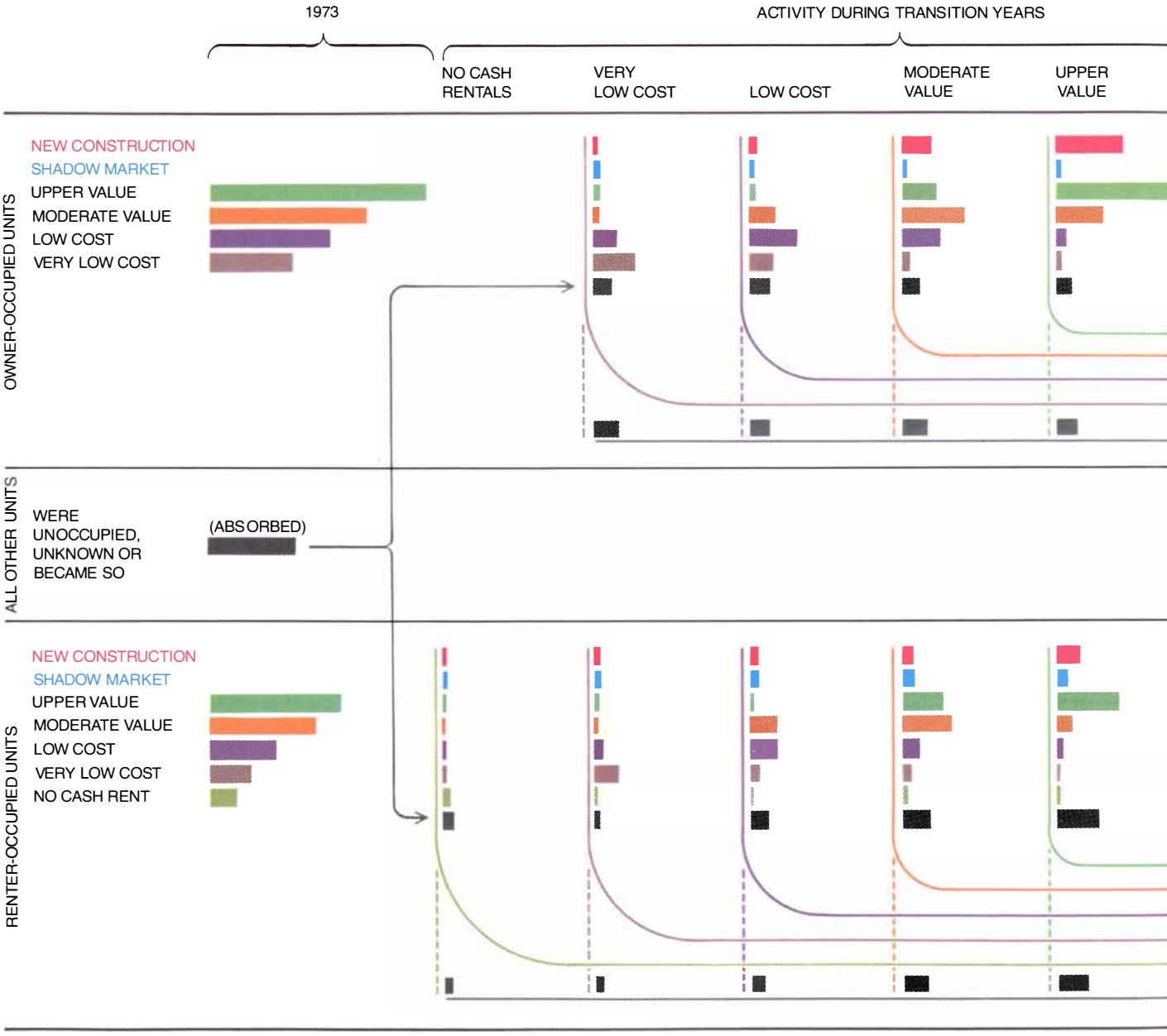


COST CATEGORIES changed between 1973 and 1980 for many housing units that were otherwise unchanged. (The U.S. Bureau of the Census calls them "same" units.) About half of the units remained in the same cost category. Of the others, more filtered down in value than filtered up, but filtering up was significant. This internal change of units was the major contributor to the number of units in a given category, outstripping new construction and the shadow market combined. It was important as a source of lower-cost rental units; in that value category other activities removed more units than they added.

many rental additions as there were owner-occupied ones during the 1973–80 period. Another difference is that as a source of additions the shadow market was about four times more important for rented units than for owner-occupied units. No doubt many of the rental units produced by the shadow market come from units previously occupied by owners. As the units age, the shadow market transforms them into a more intensive classification, usually rentals.

As for cost categories, the differences between new construction and the shadow market were even larger. For example, the shadow market was inconsequential (3 percent) as a source of owner-occupied units in general; even for the low category it contributed less than 20 percent. For the very low category, however, the shadow market accounted for 62 percent of owner-occupied additions. For rentals the impact of the shadow market was even more pronounced. It contributed 17 percent of the additions in the upper category and about 65 percent for the low category, dropping to 49 percent for the very low group. (The decline was due to Federal subsidies that encouraged new construction in the very low category, diminishing the role of the shadow market.)

The net difference between additions and removals reveals some surprises and an explanation for the growing problem of affordability. Although additions far surpassed remov-



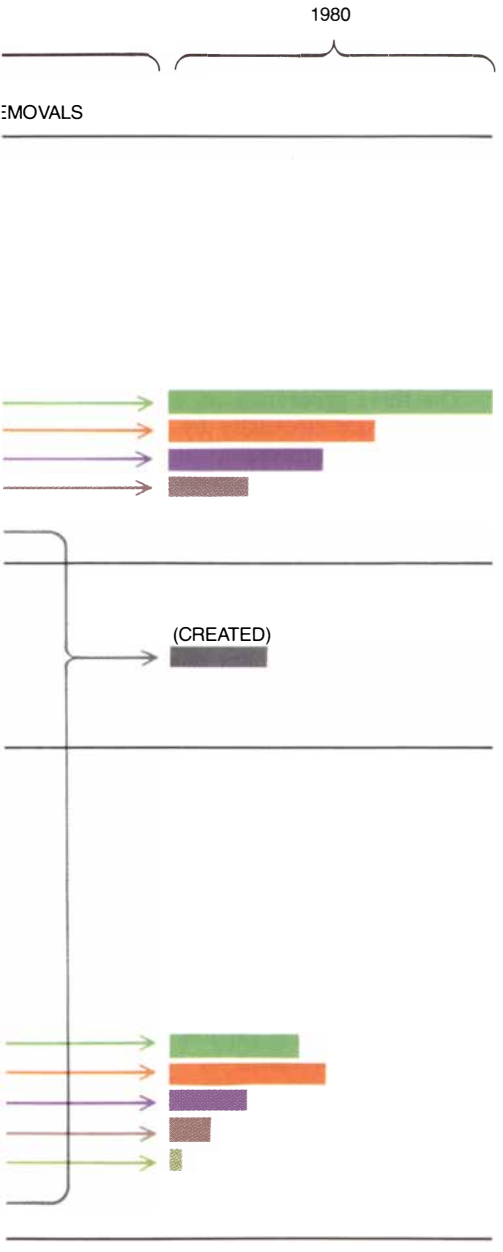
FLOW DIAGRAM shows the sources and disposition of the nation's housing stock from 1973 to 1980. It also reflects the impact of different types of change on the value categories. At the left in the diagram are the units that existed in 1973. The bars there and elsewhere are scaled in millions of units. The center of the diagram shows internal and external changes during the seven years. The right side of the diagram shows the net result of those changes as of 1980. Reading "Upper value" horizontally, one sees that by 1980 some of the 1973 units had dropped into lower-value categories and a few had been

als in the aggregate, they did not do so for every group. In the lowest two rental categories removals exceeded additions. Therefore low-income households, mostly renting, had to seek low-cost housing in the fraction of the stock that suffered a 20 percent loss from removals. The problem of affordability was particularly acute for them.

In sum, the shadow market is a strong producer of rental units. It is also good at producing low-cost units for both owner-occupants and renters. Nevertheless, neither the shadow market nor new construction fully explains the number or the source of units that existed as dwellings from one period into the next. How is it that although removals exceeded additions among low-cost rental units, a count of very low units in 1980 showed an increase over 1973? The additional units came from the existing stock, chiefly from housing that was in the high-cost group in 1973 but declined in value over the intervening years. This explanation is oversimplified, but it sets the stage for examining changes affecting units that were neither added nor removed during the seven-year period but instead remained approximately the same. Indeed, the Census Bureau calls them "same" units.

Those units call for a closer look. Although they remained nominally the same, many of them underwent some physical change during the seven-year period. Even if they did not, many of them moved to a different category of value or rent. The result was a significant impact on the number of units in each category in 1980.

One's intuition is that the movement of those units is downward in value. Indeed, casual discussions of the housing market often refer to "filtering" as a source of housing that becomes available to lower-income households. The CINCH data confirm the existence of this downward movement, but they also show that a significant number of units move upward. For owner-occupied units the figures were 18 percent downward in value and 26 percent upward; among rental units the movement was 36 percent downward in rent and 16 percent upward.



removed. Reading downward from the other "Upper value" label, one sees that new construction was a more important contributor to the 1980 stock than the shadow market was and that some of the upper-value units of 1980 came from housing that was in categories of lower value in 1973.

What were the results of all these changes taken together? The overall process can be represented by a flow diagram that starts with the 1973 stock of housing and shows the additions, removals and shifts in "same" units that affected it [*see illustration on these two pages*]. The diagram serves three main purposes: it shows better what underlies the aggregate data, it demonstrates the importance of the existing stock as a major source of lower-income housing and it reveals the importance of the shadow market.

One finding is that even though removals were much fewer than additions, they reduced the amount of housing in the low and very low categories far out of proportion to their impact on the total stock. A second finding, which I alluded to above, is that new construction contributed relatively little to the supply of low-cost housing, and much of what it did contribute was subsidized. Third, the dynamics of the housing market greatly favor owner-occupied over renter-occupied units; many fewer of the former were removed and many more were added. Since rental units are a major source of low-income housing, this finding suggests that public policy ought to give more attention to them. Finally, the shadow market is a strong contributor to low-cost housing for both owner- and renter-occupied units. Its processes must be better harnessed to public-policy goals if the affordability problem is to be dealt with successfully.

How can that be done? The answer lies mainly at the local level. Housing policies of all kinds take physical form only after they have been processed through local building codes, zoning ordinances and other regulations. Such regulations are often viewed less as housing policies than as land-use policies or as rules protecting the public's health, safety and welfare. Nevertheless, the regulations can encourage or discourage the processes of the shadow market.

At present local regulations tend to inhibit the shadow market. A case in point is their continuing emphasis on single-family housing at a time when most households are of more varied composition: female or male heads of household with no spouse present, couples of the same sex or simply unrelated people living together. Often the people in these households are elderly, and often (whether the occupants are elderly or not) the households are poor. They require a wider variety of dwellings than single-family housing or even new apartments, and in any event they need less expensive housing than can be provided by new construction.

In many areas older single-family housing in the inner suburbs could be converted to serve those needs. Even when the elderly owners are still there, they are often on fixed incomes and would welcome the additional income that would result if the single-family house could be converted into a two-family house or to provide apartments, flats or rooms for boarders. The shadow market could also encourage the use of older and currently vacant or underutilized commercial and industrial space.

All too often the local regulations now in place thwart or impede such modifications. In so doing they unnecessarily bar a fraction of the population from a ready-made source of low-cost housing. To the extent that Federal funds for housing continue to be reduced, this deprivation will be made even more severe.

Engineering Voyager 2's Encounter with Uranus

Difficult problems posed by vast distances, low light levels, aging equipment and mechanical breakdowns were solved by radio control from the ground as the Voyager 2 *spacecraft hurtled toward Uranus*

by Richard P. Laeser, William I. McLaughlin and Donna M. Wolff

On January 24 of this year the *Voyager 2* spacecraft came within 81,000 kilometers of the cloud tops of Uranus. From that unprecedented vantage point the craft was able to transmit spectacular images of the planet, its moons and its rings across the three billion kilometers that separate it from the earth. In addition to acquiring the striking visual images, the assortment of scientific instruments on the spacecraft collected copious amounts of other valuable data on the Uranian system—enough data to keep planetary scientists occupied for years to come. What made this remarkable feat ultimately possible was more than resolve and a measure of good luck. The necessary additional ingredient was extensive in-flight engineering: the monitoring, control and modification of *Voyager 2* during the course of its travel through interplanetary space.

Indeed, the very decision to send *Voyager 2* to Uranus was made three and a half years after the craft's launch into space in 1977. *Voyager 2,* like its twin *Voyager 1,* had been designed only to explore Jupiter and Saturn. By January, 1981, when the decision to add Uranus to the itinerary of *Voyager 2* was made, the two spacecraft had completed all the combined mission objectives. *Voyager 1* had encountered Jupiter and Saturn, returning a wealth of information about the two planets, including evidence of active volcanoes on Io, one of the Jovian satellites. *Voyager 2* (which was actually launched before *Voyager 1*) had arrived at its closest approach to Jupiter and was on its way to its follow-up exploration of the Saturnian system.

Yet the Voyager-project team had a compelling reason to extend the mission: the alignment of the outer planets would allow *Voyager 2* to be flung in the direction of Uranus and Neptune by the gravitational force of Saturn. (Such a gravitational "slingshot" had in fact been exploited to send the Voyagers from Jupiter to Saturn.) The last time the outer planets were similarly aligned was during the presidency of Thomas Jefferson about 180 years ago. The uniqueness of the opportunity outweighed the fact that the probability of *Voyager 2*'s lasting for another five years was estimated to be between 60 and 70 percent, well below the National Aeronautics and Space Administration's usual criterion for approving such missions. And so the decision was made to go for Uranus.

As it passed Saturn *Voyager 2* was 1.5 billion kilometers from the earth and about four years old. When it arrived at Uranus, it was twice as far away and twice as old. Yet it had in fact become better: most of the spacecraft's subsystems had actually been improved en route. The project engineers had accomplished this in spite of the difficulties posed by the vast distances, the spacecraft's age, the low light levels and a dwindling power supply. The result of their efforts was the nearly flawless encounter with Uranus, with more to come: the prospect of a successful encounter with Neptune in 1989.

The Voyager probes are a far cry from the sleek spacecraft often depicted in motion pictures and television. The dominating feature of the Voyager is its large parabolic-dish antenna, 3.7 meters in diameter, through which it transmits and receives radio signals to and from the earth. Extending in roughly opposite directions from the base of the antenna dish are two trusses. One of them supports three radioisotope-thermoelectric generators and the other supports a platform on which an array of scientific instruments, including the imaging system, are fastened. Between the two trusses are other ungainly extensions: a long boom that holds a low- and a high-field magnetometer, as well as two wire antennas for plasma-wave and radio-astronomy observations.

The three generators on the spacecraft produce electricity by applying heat, which is released by the radioactive decay of plutonium oxide, to a thermoelectric material. Such a material converts heat directly into electricity. The efficiency of the generators is low (about 5 percent); their great advantage is that they can operate in the dark outer regions of the solar system, where photovoltaic cells (which convert light directly into electricity) would be ineffective.

Along with the two cameras of the imaging system, three devices are housed on the platform opposite the generators: a photopolarimeter, an instrument that measures the intensity and angle of polarization of light, and two spectrometers, which record the component wavelengths of electromagnetic radiation. One spectrometer operates in the infrared region of the spectrum and the other in the ultraviolet. The platform is capable of scanning in two directions by means of gear-driven actuators: azimuth (side to side) and elevation (up and down). Attached to the scan-platform truss is an assortment of elementary-particle detectors. All this equipment, in addition to the magnetometers and the wire antennas, enables the Voyager spacecraft to perform 10 different scientific experiments. The spacecraft's radio system, which is its vital link to the earth, is also the means by which an 11th, radio-science experiment can be carried out.

The Voyagers' most important on-

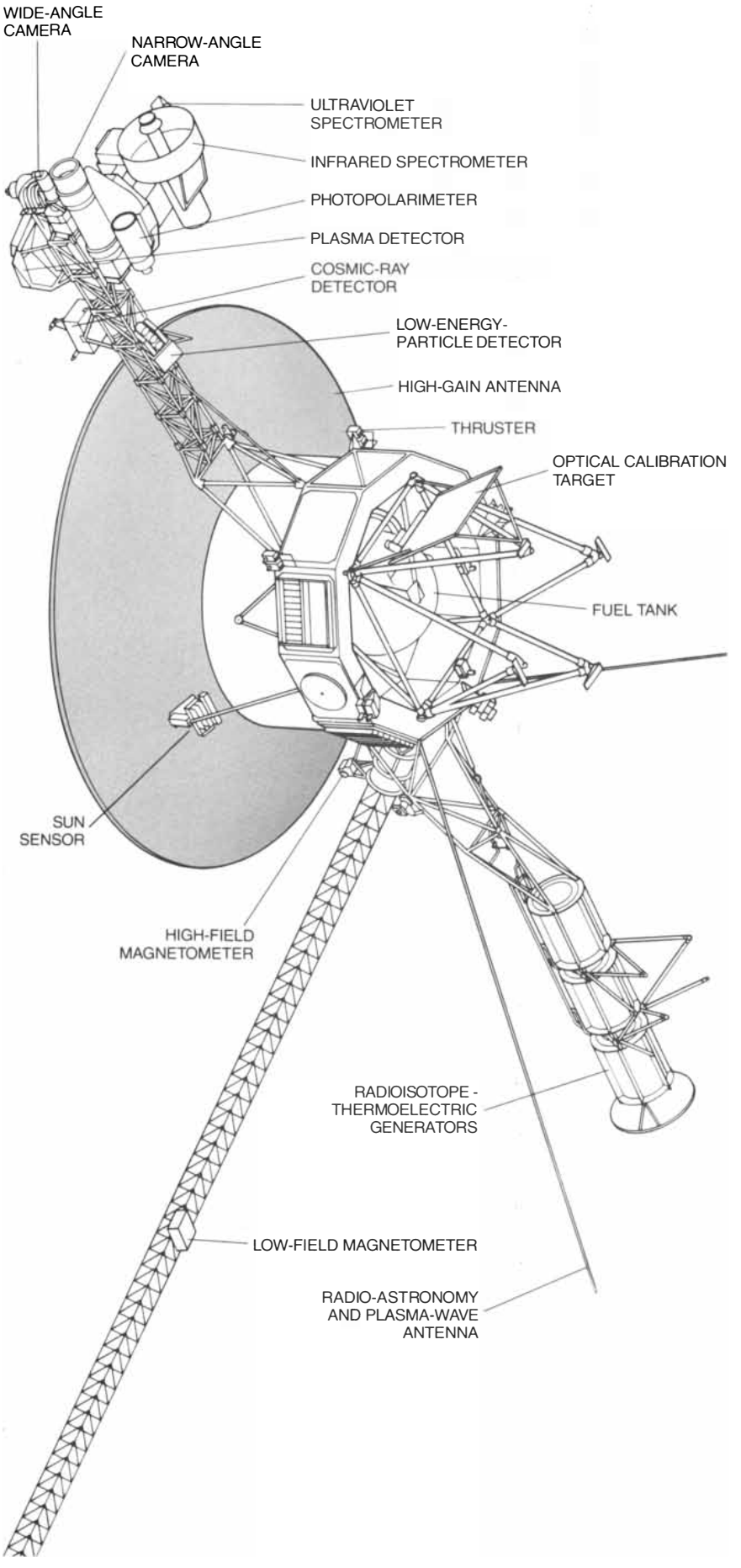
board data-handling equipment is the Flight Data Subsystem (FDS), which consists of two computers (one serving as a backup for the other). These computers, among other tasks, control the state of the scientific instruments and put the data obtained from them into the correct format for transmission to the earth.

Two different sets of thrusters control the orientation and navigation of the spacecraft. The thrusters expel gases produced in a catalyzed reaction of hydrazine, a compound of hydrogen and nitrogen that is stored as a liquid in a tank. The on-board supply of hydrazine is critical since it fuels the attitude-control thrusters that keep the highly directional parabolic antenna pointed toward the earth. The Voyager spacecraft are essentially projectiles in free flight; their speed and direction are affected by gravitational interaction with solar-system bodies, by small and continuously decreasing forces that photons emitted by the sun and the spacecraft's own power source exert and by their own thrusting. Careful management of the hydrazine left the *Voyager 2* fuel tank about half full after the Uranus encounter.

The hydrazine fuel is just one of several items that are unavoidably depleted during the course of the Voyager mission. Others include the decaying radioisotope in the thermoelectric generators and, more generally, various parts that are subjected to wear and tear. The only way to minimize the loss of these resources and thereby extend the effective lifetime of the spacecraft is to use them sparingly.

The electric-power supply actually proved to be a bigger problem than the fuel supply during the Uranus encounter. When *Voyager 2* was launched, the thermoelectric generators had a power output of more than 470 watts. By the time the spacecraft reached Uranus the natural decay of the plutonium oxide had reduced the available power output to about 400 watts, which was not enough to operate all spacecraft subsystems simultaneously. As a consequence some subsystems could only be switched on after others had been switched off, and this put certain con-

VOYAGER MOCKUP, along with duplicates of the spacecraft's various subsystems, is available at the Jet Propulsion Laboratory for testing new procedures that might be tried on *Voyager 2* as it moves toward Neptune following its successful encounter with Uranus this past January. The mockup differs from the actual spacecraft in that the magnetometer boom (*see illustration on next page*) is not extended.



straints on the design of the mission.

As the spacecraft rounded Uranus, for example, it was eclipsed by the planet's rings and the planet itself, an event called an occultation. The hope was to gather information on the temperature and composition of the planet's atmosphere and the size of the objects that form the rings by noting how radio signals beamed to the earth from the spacecraft were affected as the signals traversed the rings and the atmosphere during *Voyager 2*'s passage behind Uranus. For this experiment the project investigators planned to switch one of the spacecraft's transmitters to high power, adding 53 watts to the electric load.

During that time the investigators also planned to make pictures of the dark side of the planet and its backlighted rings. These activities involved pointing the cameras by moving the scan platform and recording the images on tape. Both activities would be additional consumers of power. If for some reason the subsystems on the spacecraft attempted to draw more than the available power, an overload-protection system would automatically turn off several subsystems. Such an occurrence in the middle of the occultation would have interrupted the scientific observations until restoration commands could be sent to the spacecraft from the earth.

Hence a careful ballet of turning subsystems and heaters on and off had to be performed in order to ensure adequate power throughout the period of the occultation experiment. The choreography was not without risk: it required operation of the spacecraft near the 400-watt limit. To minimize the risk a series of tests in which the spacecraft was operated close to the limit of available power was run on *Voyager 2* prior to the occultation. The results suggested the procedure would indeed work.

Another important resource in short supply near Uranus was one not carried on *Voyager 2:* light. Although Uranus is twice as far from the sun as Saturn is, the level of the light at Uranus is only a fourth of what it is at Saturn. That meant a fourfold increase in exposure times for the cameras, which

ELEVEN EXPERIMENTS can be carried out by various remote-controlled scientific instruments on *Voyager 2*. The spacecraft is also outfitted with an electric-power supply, communication equipment and data-handling computers. Two sets of thrusters serve to control the orientation of the spacecraft as well as its flight trajectory.

in turn increased the likelihood of blurred pictures from unintentional jiggling of the spacecraft while the camera shutters were open.

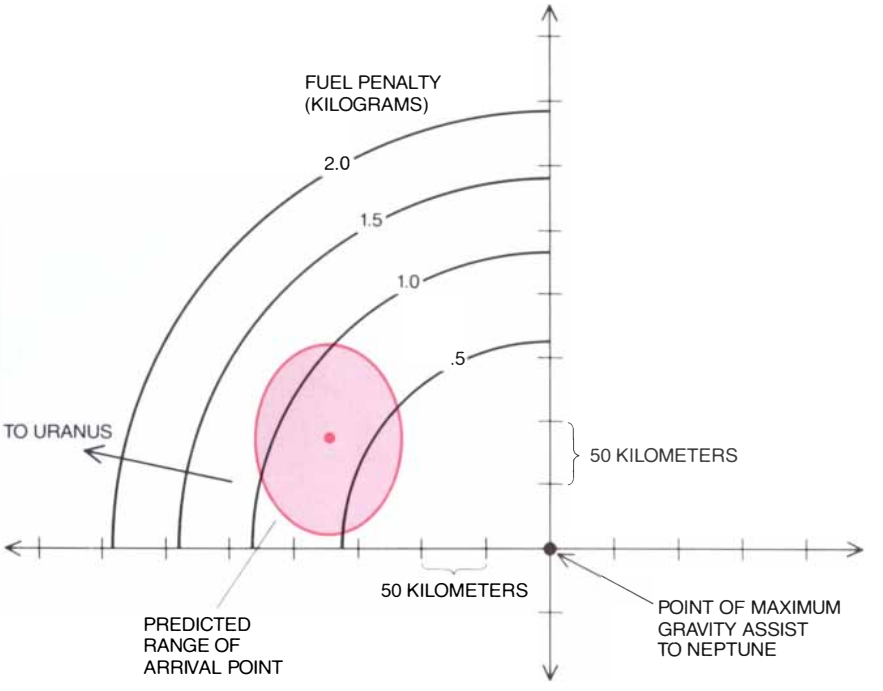
The solution to that problem was achieved by two engineering changes designed to steady the spacecraft as an observing platform. The first change addressed the unwanted angular momentum imparted to the spacecraft every time its tape recorder was switched to the high-speed mode employed to store digitized images and other data. To cancel that momentum the logic coded in the on-board attitude-control computers was modified to briefly fire the appropriate thrusters whenever the recorder's high-speed mode was switched on or off.

The second change required a still more delicate corrective use of the attitude-control subsystem. Two celestial sensors, one pointed at the sun and the other at a reference star, provide the information on the actual orientation of the spacecraft. When the spacecraft drifts too far from the desired orientation, the appropriate pitch, yaw or roll thruster is fired briefly to nudge the vehicle back to its correct position. Drifting, such as that caused by the residual torques from the tape recorder or by external torques from the pressure of solar photons, is controlled to within .05 degree along the three rotational axes during planetary-encounter periods. This range of maximum allowable drift is called the dead band.

Clearly, image smear would be reduced if drift rates through the dead band could be reduced. The Voyager engineering team analyzed the control logic and found a way to reprogram the attitude-control computer in order to reduce the length of time the thrusters are fired once the spacecraft has reached the limit of the dead band. As a result the force of the thrusts could be reduced and shifts in the spacecraft's orientation would be gentler.

Although it is simple in principle, this procedure raised practical difficulties. Ground testing on duplicate thrusters was done to determine how far the thrust levels could be reduced and still maintain firm control of the spacecraft's orientation. A reduction by a factor of about two in the thrust appeared to be feasible. Additional ground testing confirmed that the reduced-thrust mode would not shorten the expected lifetime of the thruster assemblies.

After the logic of the control algorithm was verified with a laboratory simulator, the next step was to incrementally reduce the thrust impulse on *Voyager 1,* observing the results at each step. Although *Voyager 1* has no more planetary encounters in its itinerary, its on-board scientific experiments still provide valuable data. Furthermore, it remains a valuable engineering tool in that it serves as a convenient test-bed for new procedures proposed for *Voyager 2*. Since the procedure appeared not to have resulted in any problems with *Voyager 1,* it was implemented on *Voyager 2*.



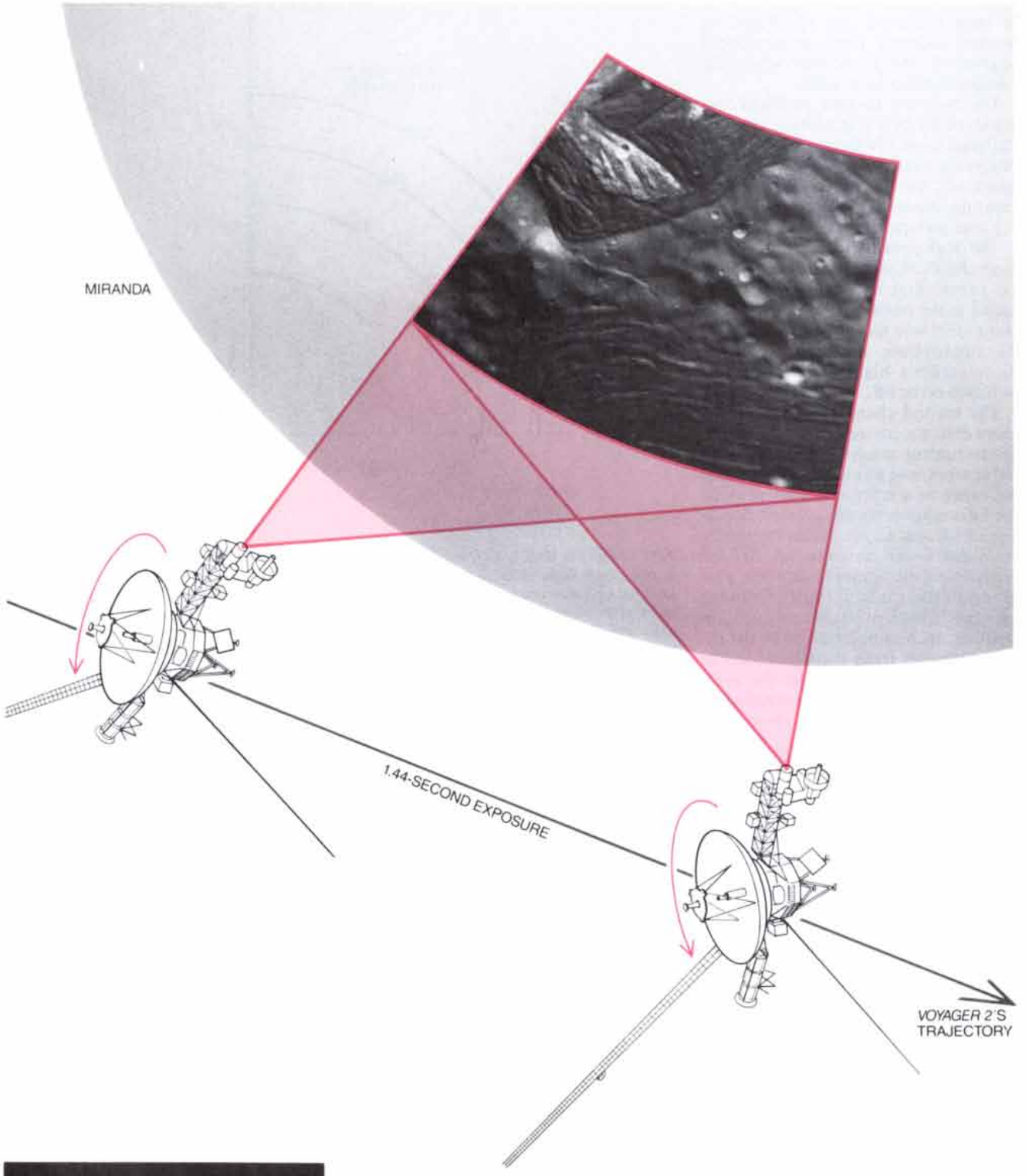
***VOYAGER 2*'S FUEL CONSUMPTION is minimized by taking full advantage of the gravitational force of large solar-system bodies to deflect the craft toward its targets. The penalty (*quarter circles*) for not reaching the point where Uranus' gravity would have most assisted it on its journey to Neptune was estimated, five days before the closest approach, to be between one-half and one kilogram of fuel. The project team elected not to perform a trajectory-correction maneuver at that time since the fuel penalty was not costly (62 kilograms of fuel remained in the tank), it would have made tracking the spacecraft more difficult and the team's limited manpower resources could be better applied to other important matters. The coordinate plane is drawn so that it passes through the center of Uranus, some 100,000 kilometers in the direction indicated, and is nearly perpendicular to the direction of travel of *Voyager 2* as it approached the planet. The fuel carried on the spacecraft is hydrazine, a compound of hydrogen and nitrogen. It is stored in liquid form but is converted into a gas before it is expelled through the thrusters.**

Although the two stabilizing procedures did help to make the images recorded at Uranus a lot sharper, *Voyager 2* was scheduled to approach so close to the planet's five known satellites that in their case another source of picture smear had to be considered: the relative motion of the spacecraft with respect to the satellites. Such a situation is familiar to any amateur photographer who has snapped pictures from a rapidly moving vehicle: the objects in the foreground are blurred even when background objects remain quite sharp.

The strategy in *Voyager 2*'s case was much the same as the one experienced photographers resort to: "panning" the camera. The camera is moved while the shutter is open in order to hold the object of interest in a fixed position within the camera's field of view.

That type of motion compensation was achieved in the case of *Voyager 2* by firing the attitude-control thrusters to turn the entire spacecraft at the proper rate while the images were being made. A similar motion-compensation procedure carried out solely by the scan platform, on which the cameras are fixed, would have been too jerky. The necessary three-axis turn rates were implemented by "fooling" the attitude-control subsystem: a message was sent to the spacecraft that commanded the attitude-control logic to compensate for a fictitious drift in the attitude-reference signal. An undesirable consequence of this process was that the spacecraft antenna would be pointed away from the earth. The temporary loss of radio communication was outweighed by the value of the high-resolution images acquired.

Like the light levels, the supportable bit rate, which is the amount of digitally coded information that can be reliably received in a given time, is inversely proportional to the square of the distance of the spacecraft from the



HIGH-RESOLUTION IMAGES OF MIRANDA, one of the moons of Uranus, are among the best images ever made in the course of a planetary flyby. Because of the low levels of sunlight in the vicinity of Uranus, long exposure times were necessary for the various images made during the encounter. To avoid blurred images, the computer logic that operates the firing of the attitude-control thrusters was rewritten to minimize jiggling of the camera while the shutter was open. More important in the case of Miranda was the precise firing of the thrusters to rotate the entire spacecraft smoothly in such a way as to compensate for the relative motion of the spacecraft and its target. During the exposures *Voyager 2* was traveling at 72,000 kilometers per hour as it sped by within 29,000 kilometers of Miranda. The clarity of the images eventually obtained made it possible to resolve surface features as small as half a kilometer. If the spacecraft had not been accurately guided, positioned and rotated, the resolution of the pictures would have been about 26 kilometers, comparable to that of the smaller picture of Miranda on the left, and many of the satellite's topographic characteristics would not have been visible.

earth. In contrast, the amount of data scientists want to collect tends to increase with distance. The way around the dilemma in the case of *Voyager 2* was to modify the data-processing software run on the spacecraft's computers and to enhance the capability of the receiving stations on the earth.

On *Voyager 2* data streams are coded prior to transmission so that errors can be detected and corrected on the ground. A simple example of such a coding scheme is one in which each bit, or digit, of a binary number (a 1 or a 0) is appended to a pair of repetitions of the bit. In other words, if the data bit were 1, 111 would be transmitted. Similarly, 000 would be transmitted for each 0 data bit.

The fact that a 111 is transmitted by the spacecraft does not guarantee that 111 will be received at a ground station. Because of unavoidable background noise in radio-signal reception and processing, occasionally a 1 will be recorded as a 0, or vice versa. Therefore three-digit binary numbers could be received as eight possible combinations of 1's and 0's. Yet if a "majority rules" decoding tactic is applied, with the data bit assumed to be the value of whatever digit is repeated at least once in the transmitted triplet, then the overall probability that the data bit will be erroneously interpreted is significantly less than it would be if the data bit were transmitted without encoding and decoding.

Although that is a simple example, it gives a sense of the tradeoffs involved in applying any error-detecting and error-correcting code: one can improve the reliability of data reception, but only by increasing bit overhead, or the total number of bits transmitted. For *Voyager 2,* however, any increase in the total number of bits that have to be transmitted reduces the rate at which scientific data can be returned.

In the example cited above two coding bits constitute what is called the coding block and one data bit what is called the data block; the coding block is twice the size of the data block, in effect tripling the transmission load. The actual coding schemes used on Voyager are naturally much more sophisticated than the example, and the bit overhead is fortunately not quite as large. During the Jupiter and Saturn encounters the particular coding scheme used, called Golay coding, had resulted in code blocks that were equal in size to the data blocks. At Uranus another coding scheme was employed (Reed-Solomon coding), so that the code block was only one-seventh the size of the data block. The decision to turn to Reed-Solomon coding at Uranus carried with it some risk. The onboard hardware for implementing it was nonredundant, whereas the Golay hardware had a backup.

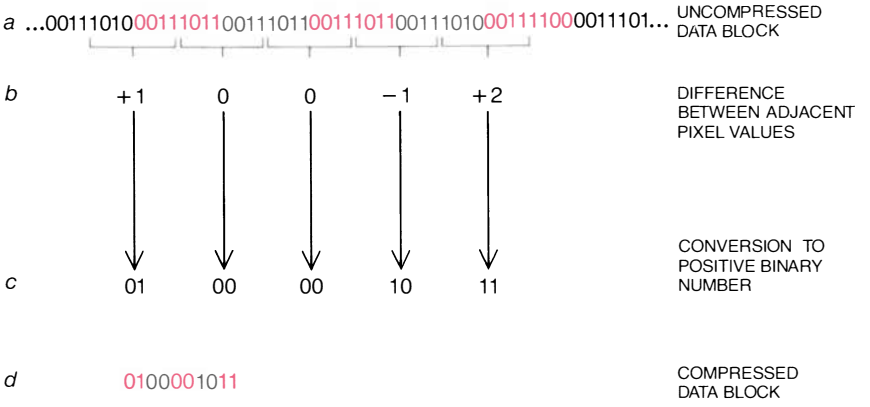


IMAGE-DATA COMPRESSION was applied during the Uranus encounter to reduce the number of bits that had to be transmitted to ground stations. Each line of an image consists of a row of 800 pixels, or picture elements. The brightness of each pixel is expressed as an eight-bit binary number. Rather than transmitting the brightness value for each pixel to the earth, as was done for the earlier Saturn and Jupiter encounters, only the difference in brightness levels between adjacent pixels was transmitted. A simplified version of the technique is illustrated. First the 800 pixels in each row are divided into 160 groups of five pixels each (*a*). The differences in brightness between the first pixel in a group and the last one in the preceding group and between adjacent pixels within the group are calculated (*b*). The differences are then converted into positive binary numbers (*c*). (In this example two-digit numbers suffice.) The final compressed data block transmitted to the earth consists of the five two-digit numbers strung together. In this example data that would have required 40 bits to express are compressed into a block of 10 bits.

Another measure that reduced still further the number of bits that had to be transmitted back to the earth from Uranus was "compression" of the image data. The cameras of *Voyager 2* decompose the image in the field of view into an array of pixels, or picture elements, much as pictures in newspapers consist of tiny dots. The spacecraft transmits to the ground a binary number that corresponds to the brightness level of each pixel.

A Voyager image is composed of 800 rows of pixels, each row consisting of 800 pixels, for a total of 640,000 pixels. Each pixel can assume one of 256 different brightness levels ranging from black to white. To express such a range of brightness levels (a total of 2^8 levels) in binary code, eight bits would be necessary. Hence it takes a total of 5,120,000 bits ($800 \times 800 \times 8$) to transmit one picture (not counting the bits that are added to serve as error-detecting and -correcting bits).

The number of bits necessary to express such an image can be more than halved by taking advantage of the fact that adjacent pixels usually have brightness levels that are close in value. This is particularly true for those pixels that do not straddle the borders delineating an object. If *Voyager 2* could be made to transmit only the change in brightness value from pixel to pixel rather than the absolute brightness value for each pixel, then on the average about three bits per pixel would suffice rather than eight. (Obviously the absolute brightness of the first pixel in each row must be transmitted as a starting point.)

This in fact was done. The backup function of the second FDS computer was sacrificed so that the device could be reprogrammed with data-compression algorithms. The image-compression approach is more vulnerable to errors, however, since each pixel-brightness value (except for the first pixel) is dependent on the value of the preceding pixel. Hence a one-bit error could affect an entire row of a compressed image rather than just a single pixel of an uncompressed image. Both compressed and uncompressed modes were therefore applied in the transmission of the almost 6,000 images obtained of the Uranian system.

Not all the measures taken to enhance the reliable return of data took place on the spacecraft. The bit-

collection capability of the ground stations was improved by electronically combining the signals received by several antennas on the earth, a procedure called arraying. For example, the 64-meter antenna dish and the two 34-meter dishes of NASA's Deep Space Network tracking complex in Australia were arrayed together and then combined with the 64-meter Parkes radio telescope, which was borrowed from Australian radioastronomers. The rate at which data could be reliably received was increased from 14.4 kilobits per second for a single 64-meter antenna to a potential 29.9 kilobits per second for the array. In practice the highest data rate at Uranus was held to 21.6 kilobits per second; the rate of 29.9 kilobits per second was reserved for emergency situations.

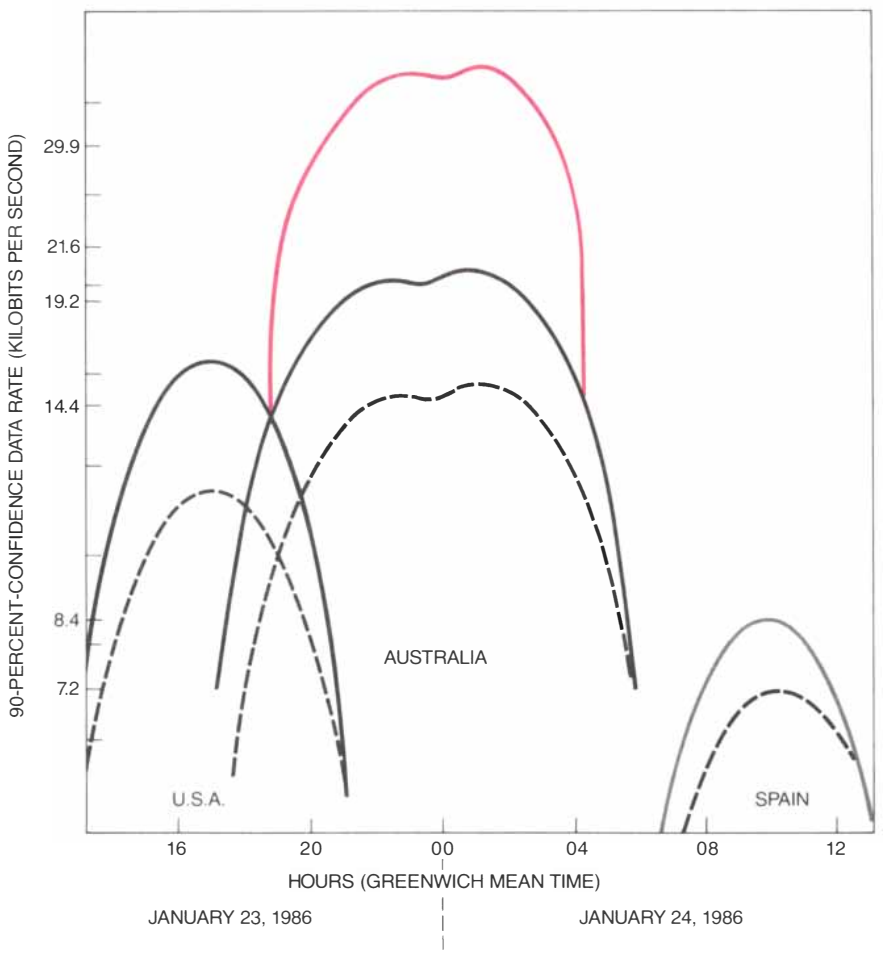
Another activity of importance that took place on the ground had to do with the steering of the spacecraft. Large, earth-based computers have to be relied on to do trajectory calculations based on sophisticated celestial-mechanics models and statistical techniques. Nevertheless, interplanetary navigation remains a mixture of science and art. To produce sensible answers a certain amount of human judgment is required.

Input to the navigational process can be classified into two basic types: radiometric and optical. Radiometric data rely on the interaction of a spacecraft's radio system with ground-based antennas in three different ways: the Doppler shift of radio signals (the shifting of the radio frequency due to the relative motion of the transmitter and the receiver), the signal's round-trip time and the angle between the spacecraft's radio beam and a reference radio-signal source in the sky. These three measurements respectively yield information about the line-of-sight component of the spacecraft's velocity, its distance and its angular location on the celestial sphere.

Optical data are produced by the two cameras on the Voyager spacecraft. Optical navigation, which was first tested on Mariner and Viking probes to Mars, was in fact indispensable in the Voyager mission. Unlike radiometric data, optical data enable one to determine the spacecraft's position in relation to the objects of interest—Uranus and its satellites in the case of *Voyager 2*—rather than in relation to the earth. The optical approach allows one to determine the orbits of natural satellites, which do not have radio receivers, as well as the orbit of the spacecraft. This is of considerable importance when one is trying to point instruments at a nearby satellite, particularly if its mass and orbit are not known precisely.

Both radiometric and optical data were processed to pin down *Voyager 2*'s location to within 23 kilometers from more than three billion kilometers away. Such precision actually was greater than that required for accurate pointing of the cameras during the close flyby of Miranda, one of the moons of Uranus, but it was not achieved without travail.

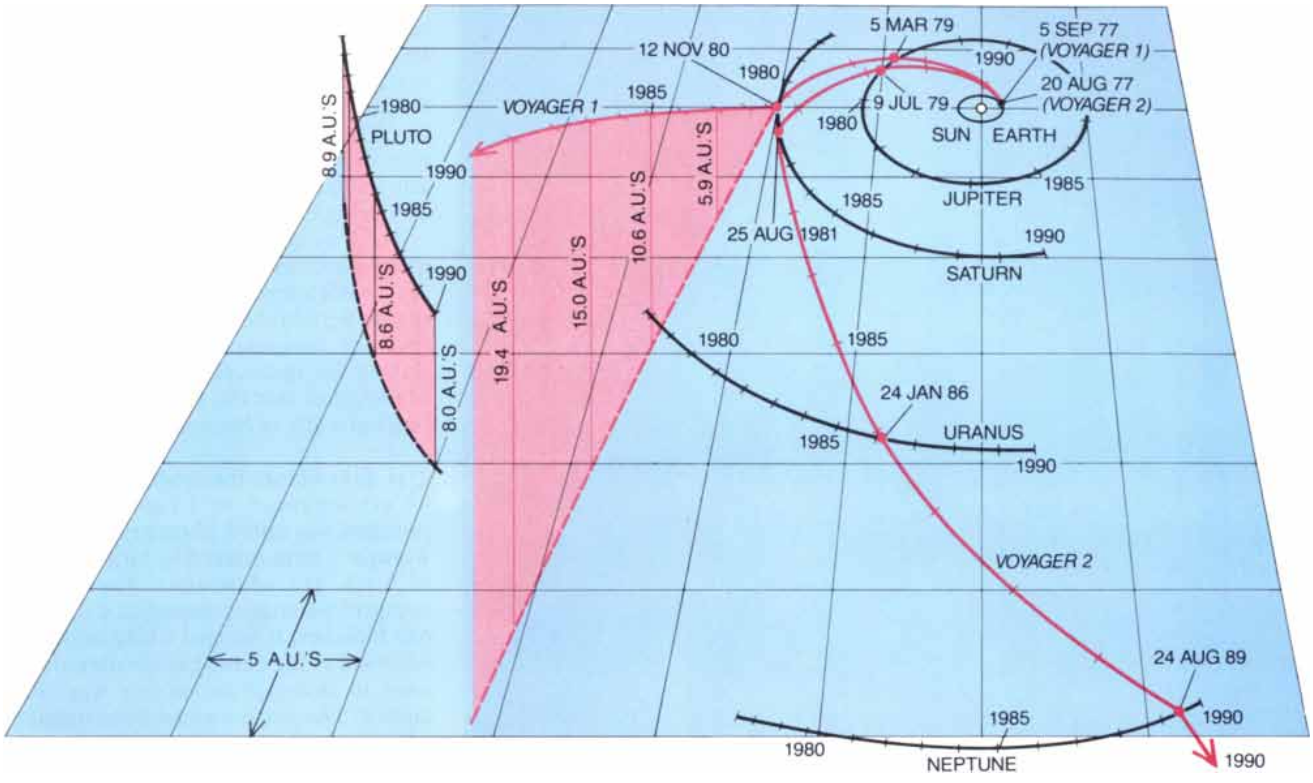
In December, 1985, some difficulties were encountered in fitting the navigational data to the dynamical model of the flight path. It took a relatively large change in the estimate of the mass of Uranus, amounting to an increase of .3 percent, before the difference between predictions and observations became small and the navigational process could go forward smoothly. It turned out that the orbit-determination programs had calculated the mass of Uranus incorrectly and needed to be reset "manually."



COMMUNICATION CAPABILITY was augmented during the *Voyager 2* Uranus encounter by electronically combining the signals received by an array of antennas. Communication capability is measured in terms of the probability that a given data rate can be received relatively free of error. Shown here are curves representing a 90 percent probability of capturing data from *Voyager 2* as it neared the point of closest approach to Uranus. Various antenna combinations were available at the Deep Space Network complexes of the National Aeronautics and Space Administration in the U.S., Australia and Spain: a 64-meter parabolic-dish antenna alone (*broken curves*); a 64-meter antenna arrayed with one 34-meter antenna (*gray curve*), and a 64-meter antenna arrayed with two 34-meter antennas (*black curves*). By adding the 64-meter Parkes radio telescope in Australia (*colored curve*) as much as 29.9 kilobits per second could be reliably received. The best reception was achieved in Australia, because Uranus was in the southern sky then.

Failures are in a certain sense expected for the Voyager space probes, which have already exceeded the lifetime for which they were designed. When a component does fail, diagnosing the problem, correcting or repairing it or, in the worst case, trying to make do without it often tests the ingenuity of the project team.

The most serious problem the flight team faced occurred in 1978, during *Voyager 2*'s passage from the earth to Jupiter: its primary radio receiver failed. To be sure, the spacecraft was equipped with a backup receiver, but unfortunately this device was also



NAVIGATION of Voyager spacecraft is accomplished by continually determining the spacecraft's position and velocity, calculating whether it will reach the target point at the correct time and precisely steering it back on course by firing thrusters. Other forces that act on the spacecraft also have to be taken into account—principally gravitational forces and the force exerted by photons emitted by the sun. The position and velocity of the spacecraft are computed from two types of data: radiometric and optical. Radiometric data are obtained by analyzing the radio signals from the spacecraft. Optical data are obtained from the spacecraft's cameras. The flight path is plotted and needed trajectory-correction maneuvers are calculated from the data by earth-based computers. *Voyager 1* flew over Saturn's south pole and was gravitationally deflected above the plane of the ecliptic, which is defined by the plane of the earth's orbit. It is now about 14 astronomical units (A.U.'s) above the plane. (One A.U., the average distance from the earth to the sun, is approximately 150 million kilometers.) *Voyager 2,* still in the ecliptic plane, is targeted to fly over the north pole of Neptune on August 24, 1989. The encounter will deflect *Voyager 2* downward, below the plane of the ecliptic.

flawed. Its bandwidth (the range of frequencies over which signals can be received) was found to be reduced to less than a thousandth of its design specification. The cause of the reduced performance of the backup receiver has been attributed to the failure of a single capacitor.

In order to get commands through the backup receiver's narrow bandwidth window, the flight team has to keep careful account of any factor that could alter the frequency of received radio signals. Most obvious is the Doppler shifting of radio signals due to the rotation of the earth and the motion of the spacecraft. The earth's rotation can shift the frequency of the uplink signal (the signal transmitted to the spacecraft) by more than 30 times the bandwidth of the crippled receiver.

The most difficult frequency-related events to characterize are those that affect the temperature of the spacecraft. Changes in the temperature can shift the center of the available bandwidth of the receiver. Indeed, if the receiver temperature changes by just a quarter of a degree Celsius, the center of its bandwidth shifts by about 100 hertz. Hence any process that might generate or remove heat has to be monitored carefully. These processes include turning the spacecraft and switching power loads.

The effect of such processes on the receiver's bandwidth is constantly gauged by frequent on-board temperature measurements and by direct testing to determine what is the best frequency for communication. Although an extensive reservoir of knowledge has been gained over the years and techniques have been developed to maintain radio command over *Voyager 2,* communication with the spacecraft can still be disrupted for a period of a few days by major thermal events.

Just after its closest approach to Saturn in 1981, *Voyager 2* suffered a mechanical breakdown. The actuator that controls the azimuthal position of the scan platform became stuck. Some scientific data about Saturn were lost in the incident, but what concerned the Voyager team more was the possibility that the platform would be inoperable for the upcoming Uranus and Neptune encounters. To ascertain the likely cause of the trouble and to develop a plan to work around it, extensive testing was done on the ground and on both Voyagers.

At the Jet Propulsion Laboratory (J.P.L.) 86 mockups of the actuator were built and tested in an attempt to infer the state of the on-board actuator and to determine the factors (such as temperature, actuator rate and applied torque) influencing the probability of jamming. An actuator in a spacecraft simulator also proved valuable in obtaining statistical data.

Engineering judgment and statistical analyses yielded three conclusions. First, the actuator would probably function at Uranus if platform motions were limited to low speed. (Because of their high scientific value, however, four medium-rate motions were allowable exceptions.) Second, the observations at the closest approach to Uranus should be protected by having a contingency computer program ready to transmit to *Voyager 2* if the actuator showed signs of jam-

ming. The program would substitute less accurate and more difficult spacecraft rolls (using the attitude thrusters) for the motion of the azimuth actuator. Third, the condition of the actuator should be monitored by occasionally reducing the applied torque and observing its performance.

The Voyager team put these recommendations into effect, and when the spacecraft approached Uranus, the actuator performed flawlessly; the contingency computer program was not sent to the spacecraft. There is every expectation that the system will function normally at Neptune as well.

Six days before the spacecraft's closest approach to Uranus another problem was noted: photographs from *Voyager 2* were marred by large blocks of black and white lines. Since only compressed images displayed the curious blotches, it seemed likely that the software in the earth-based computers used to decompress images was the culprit. The software had been updated recently, and perhaps some "bugs" had been inadvertently introduced. Engineers of the Voyager team undertook the tedious task of decompressing selected lines of the images by hand, but the same perplexing black and white lines showed up. The only remaining possibility was a problem on the spacecraft itself.

The next day commands were sent to the spacecraft to transmit the contents of the FDS-computer memory that held the instructions responsible for compressing the images. After the readout was received at the J.P.L., what was in the computer's memory was compared bit by bit with what should have been there. It was discovered that a single bit of an instruction word, which should have been a 0, was a 1. There were two possible explanations for the incorrect bit: either a cosmic ray had caused the bit in the memory cell to flip from 0 to 1, in which case it could easily be reset to 0, or there had been a permanent hardware failure in the memory.

The FDS experts at the J.P.L. were immediately directed to write a program that could act as a patch, circumventing the possibly failed memory lo-

URANUS would have appeared as a pale, greenish blue sphere (*top*) to a human traveler on *Voyager 2*. The image is a composite made from images that were recorded with various color sensitivities. By enhancing the differences in brightness among the images, a false-color image (*bottom*) brings out atmospheric features that would be imperceptible to the unaided eye. The ring-like features seen in the images are artifacts caused by dust in the camera optics.

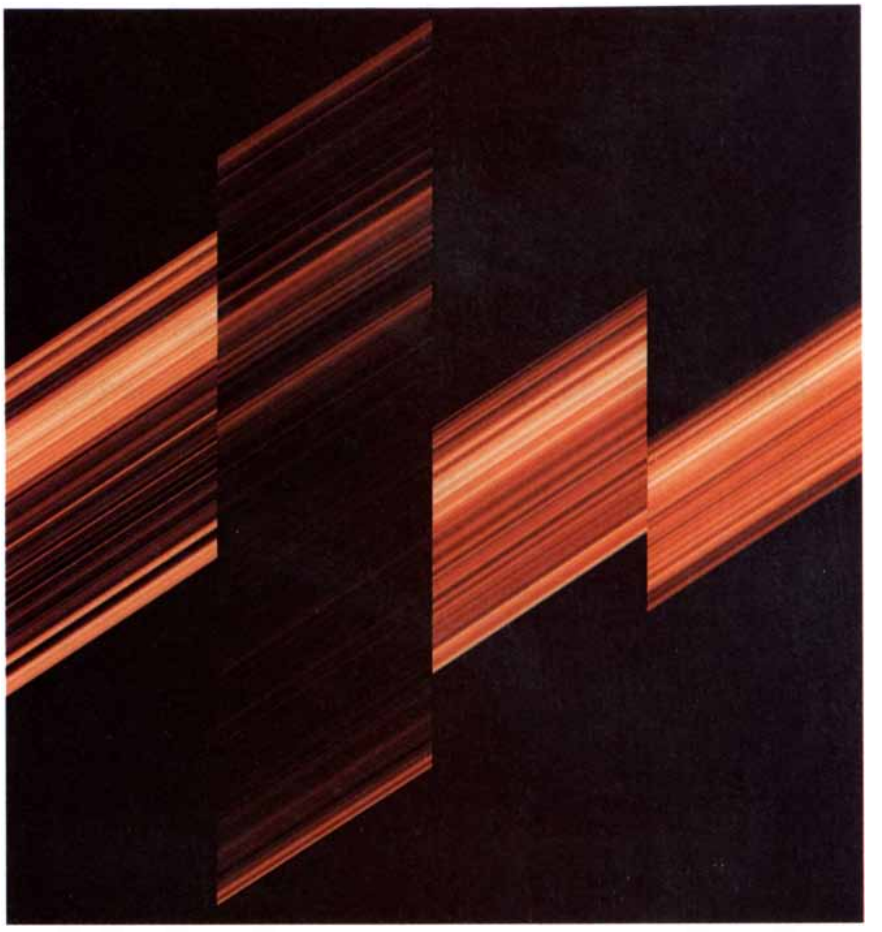
cation. The patch was programmed into the spacecraft's computer on the evening of January 20, four days before the closest approach. The next morning the transmission of fault-free compressed photographs resumed. Although the troublesome memory location had been bypassed by the patch, a command that could reset the incorrect bit was also transmitted to the spacecraft. The bit did not reset, however, and is considered to have failed permanently.

The success of the Voyager team's efforts can be measured by the mission's scientific achievements during the Uranus encounter. The improvement in spacecraft steadiness, combined with the capability afforded by data-compression and data-coding changes, facilitated the search for new objects by the imaging system and the photopolarimeter and led to the discovery of 10 new Uranian satellites and several new rings.

The accurate navigation and the motion-compensation maneuver resulted in pictures of Miranda that are among the best images ever returned by a planetary flyby mission. The execution of the power-management procedure during occultation allowed observations to proceed without a hitch, and detailed information on Uranus' atmospheric structure, including temperature and composition, was obtained. The finding that the temperatures at the poles and at the equator were nearly the same was unexpected. During the occultation one 96-second exposure of the back-lighted ring system was successful, and it revealed an intricate dust structure within the system of rings. The magnetometers on the spacecraft revealed another oddity: Uranus' magnetic field. It is tilted 60 degrees from the rotational axis. Hence as Uranus rotates with a period of 17.25 hours (a fact indicated by Voyager data), the entire magnetic field wobbles.

Barring any catastrophic failure of a vital component, a flood of valuable scientific data is also expected when *Voyager 2* encounters Neptune from June 5 through October 2, 1989—12 years after its launch. Current plans call for the spacecraft to pass over the north pole of Neptune within a few thousand kilometers of its cloud tops, making it the closest approach to any celestial body by a Voyager spacecraft. Its subsequent flight path will carry it south of the ecliptic (the orbital plane of the earth), perhaps eventually to reach the heliopause, the boundary between the sun-influenced region of space and interstellar space.

The heliopause has not been reliably detected by remote-sensing techniques; when and where it will be crossed by either of the two Voyager spacecraft is not predictable. Estimates of its distance from the sun range from 50 to 100 astronomical units. (One A.U. is the average distance from the earth to the sun, about 150 million kilometers.) If the heliopause is found at the low end of the distance estimates, the Voyagers' scientific instruments may confirm that fact in the mid-1990's, when the two spacecraft will be 50 A.U.'s from the sun.



EPSILON RING OF URANUS, outermost of the many rings that encircle the planet, has a nonuniform mass distribution and radial width. The four "slices" of the ring shown here have been reconstructed from data obtained by *Voyager 2*'s polarimeter. Reddish areas represent parts of the ring containing less material; yellow areas contain more material. The varying width of the ring is evident from the varying length of each slice, although the pair of slices at the left differ in scale from the pair at the right. The smallest features discernible are about 270 meters across. The color scheme enhances the visibility of small-scale structures; it does not show the ring's true color, which is gray.

If a fatal failure does not end the Voyagers' mission, exhaustion of one of the consumables eventually will. The rate of depletion of the most obvious consumable, the attitude-control fuel, depends heavily on interaction with other elements of the mission design. Based on the amount of fuel expended so far, there would appear to be enough hydrazine on both spacecraft to last through about the year 2030. A more serious limitation is electric power. Generator output is projected to decay to the threshold (245-watt) level in about 2013. Below this power level none of the scientific experiments can be supported. The limitation that might effectively end the two Voyagers' scientific missions is the sensitivity of the sun sensor. The amount of impinging sunlight may fall below the sensor's threshold level before the power gives out. Without a sensor to control the pointing of the antenna, the Voyagers' ability to communicate with the earth would be lost.

With continuing support from the ground, NASA's most productive scientific facility is likely to continue to provide data into the next century. The unqualified success of the Voyager project is a testament to the project team's expertise, ingenuity and determination in engineering across billions of kilometers of space.

Studying the Earth by Very-Long-Baseline Interferometry

In which radio signals emitted by quasars billions of light-years away serve as benchmarks for measuring the earth's wobble, tiny changes in its spin rate and the imperceptible drift of its plates

by William E. Carter and Douglas S. Robertson

What we propose here may at first seem fanciful: By carefully observing the faint radio signals from quasars, quasi-stellar objects that lie billions of light-years away at the edge of the known universe, one can learn important things about the structure of the earth. And yet that is just what we and other geodetic astronomers have been doing for the past decade by means of a technique called very-long-baseline interferometry (VLBI).

In VLBI radio telescopes hundreds or even thousands of kilometers apart simultaneously track the same radio source. Radio astronomers use such observations to generate images of the source that are as sharp as the ones they would get with a single continent-size antenna. Geodetic astronomers, on the other hand, extract a different kind of information from the data. In particular, we measure minute changes in the length and orientation of the baselines between the radio telescopes. When the earth's mantle and crust wobble a few centimeters over its spin axis, or when the orientation of the spin axis in space shifts by a millisecond of arc, we detect it; when the day gets a millisecond shorter or longer because the earth is spinning faster or slower, we measure the change.

The significance of such measurements is that they help theoreticians to develop improved models of the earth's internal structure and of the motions of its crustal plates. Essentially all that is known about the earth's interior has been inferred from such indirect measurements; the deepest drill holes have penetrated no more than 15 kilometers into the crust. Although today seismology is the best-known source of geophysical information, historically geodetic astronomy has been of comparable importance. The size of the earth, its ellipsoidal shape, the fact that it does not rotate at a uniform rate, that its mantle and crust wobble about its spin axis and that the axis precesses and nutates in space—all this knowledge came from centuries of observing the sun and stars, and it all contributed to current understanding of the earth's structure.

In recent decades, however, geodetic astronomy had virtually stalled. The optical techniques that relied on visible stars as a reference frame had run up against two fundamental limits. First, because the atmosphere refracts the light from stars in an unpredictable way, a star's position can be determined only to within a fraction of a second of arc. Second, the visible stars exhibit proper motion, that is, their positions change with time as they move across our line of sight. Different stars move in different directions and at different rates. Measuring the exact position of a point on the earth by referring to the visible stars is rather like being on a boat and trying to gauge one's position by observing the positions of other boats in a large, disorderly fleet. What one needs instead are fixed beacons on the shore.

Quasars are the beacons. With VLBI atmospheric refraction is much less of a problem; the position of a quasar can be determined to within a fraction of a millisecond of arc. At the same time quasars are so far away that even with VLBI no proper motion has ever been observed for any of them. Hence they provide the best available approximation to a fixed reference frame. It is also a convenient reference frame: quasars are distributed over the entire sky, and since their radio signals are not blocked by clouds, they can be observed rain or shine, day or night.

With VLBI one can detect changes as small as a centimeter in the relative positions of points on the earth. As a result the earth's wobble, the precession and nutation of its axis and the variations in its rotation rate can now be measured with unprecedented accuracy. Equally important, the sensitivity of VLBI has made it possible to measure another variable for the first time: the slow movement of the earth's tectonic plates. It is the collisions of the plates that, over millions of years, raise mountain ranges, and it is their divergence that creates oceans. Previously these motions could only be inferred from geologic evidence. Now we are beginning to measure them directly; to observe, for example, that the baselines between radio telescopes in Europe and North America are getting longer by about a centimeter per year as the continents spread and the Atlantic widens. With VLBI we are watching geology happen.

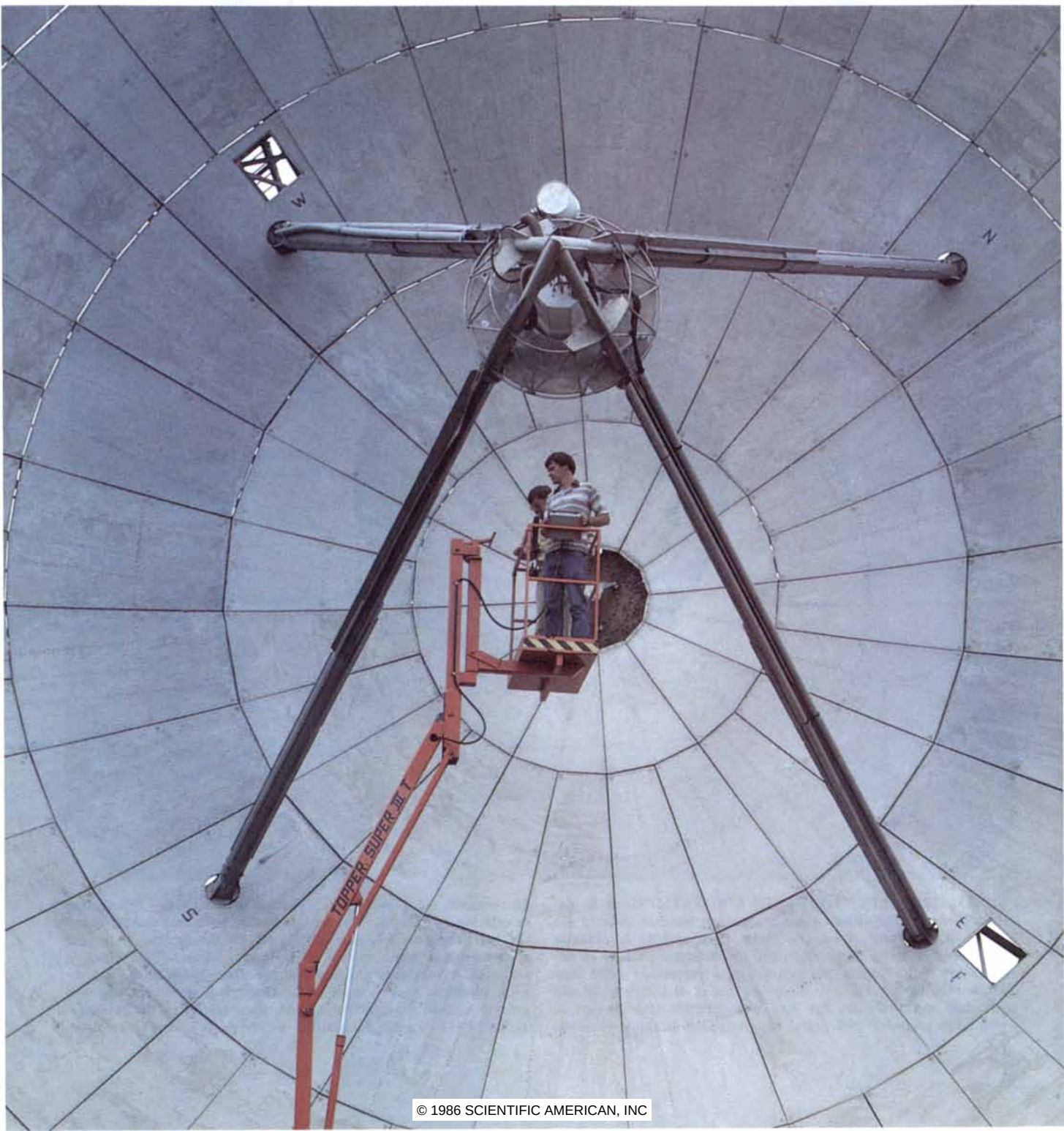
Among the earliest workers to recognize the potential for applying VLBI to geophysics were Irwin I. Sha-

RADIO TELESCOPE at the U.S. Naval Observatory facility in Richmond, Fla., is one of several with which the authors detect the faint signals from quasars. The dish is 17 meters in diameter; its panels have been precisely shaped and aligned to form a nearly perfect paraboloid. Incoming radio waves are reflected off the dish and detected by a receiver mounted on the quadripod at the prime focus of the paraboloid. In very-long-baseline interferometry (VLBI) the same quasars are observed simultaneously by telescopes thousands of kilometers apart, and the baselines between telescopes are measured.

piro, Alan E. E. Rogers and their colleagues at the Massachusetts Institute of Technology and the National Aeronautics and Space Administration's Goddard Space Flight Center. Based on their early results, the National Geodetic Survey undertook in 1977 to develop a set of three VLBI observatories exclusively dedicated to geodetic measurements. NASA and the U.S. Naval Observatory soon joined the project, which was called POLARIS, for Polar-Motion Analysis by Radio Interferometric Surveying.

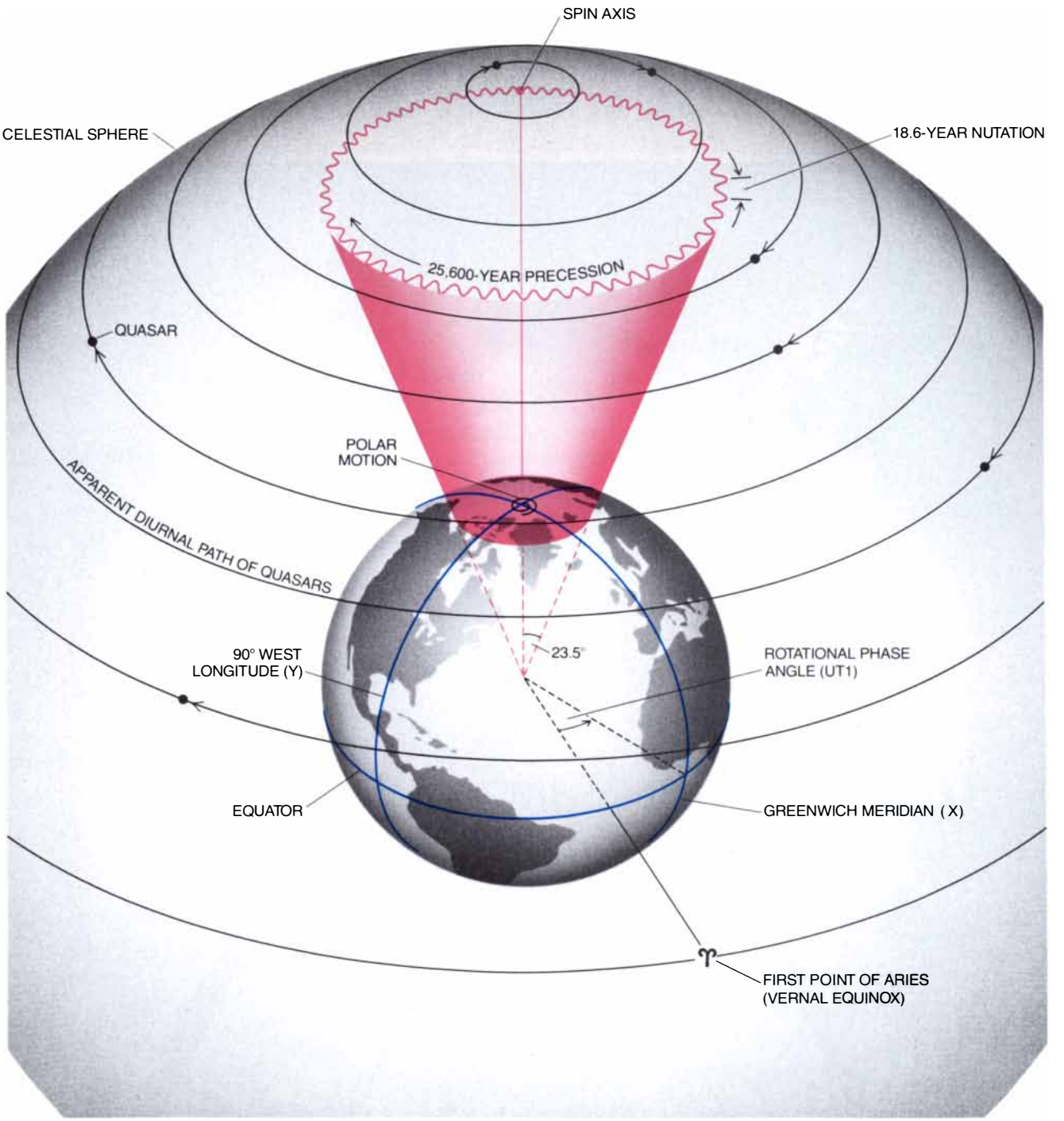
The first POLARIS station, situated near Fort Davis, Tex., and managed by Harvard University, became operational in September, 1980, and immediately began regular observing sessions in collaboration with the Onsala Space Observatory in Sweden and the Haystack Observatory of M.I.T. In June of the following year a dedicated POLARIS telescope began operating at the Haystack complex in Westford, Mass. The third POLARIS station, at the Naval Observatory facility near Miami, Fla., came on-line in January, 1984. Since then the three stations have carried out hundreds of observing sessions together and in cooperation with other observatories in the U.S., Sweden, West Germany, Japan and South Africa.

During each of these sessions two or more telescopes track the same quasars. The signals received are digitized and recorded on magnetic tapes, along with "tags" giving their times of arrival on a local time scale defined by an atomic clock. Because the arrival times are recorded independently at



each observatory, there is no need for direct communication links between the observatories. The observatories ship the tapes to a correlator center, where they are replayed and the data streams from pairs of observatories are compared.

Although quasars are extremely energetic objects (precisely what they are is a matter of ongoing controversy), their signals are extraordinarily faint by the time they reach the earth. Most of the digital bits recorded on the tapes are noise either from the receiver or from other natural and manmade radio sources. Nevertheless, one can extract a signal from the noise by comparing the bit streams from different telescopes; the noise is different at each telescope, but the signal is the same at all of them.



FLUCTUATIONS IN THE EARTH'S ROTATION can be detected through VLBI observations of quasars; because quasars are billions of light-years from the earth, they provide a virtually fixed reference frame. Precession and nutation are changes in the orientation of the earth's spin axis in space: precession is the slow motion that causes the axis to sweep out a cone every 25,600 years, and nutations are the tiny irregularities (the largest of which has a period of 18.6 years) superimposed on the precession. In contrast, polar motion is a change in the earth's orientation in relation to the spin axis. As the mantle and crust wobble about the axis, the pole traces a spiral path on the surface. VLBI also allows precise measurements of Universal Time (UT1), which is the earth's rotational phase angle: the angle between the Greenwich meridian and a great circle through a celestial reference point (the first point of Aries) and the celestial poles. Measurements of UT1 reveal fluctuations in the earth's rate of rotation.

The signal arrives at the earth as a series of planar wavefronts. By correlating the data streams from two telescopes we can determine the time delay between the arrival of a particular wavefront at the first telescope and its arrival at the second telescope [*see illustration on next page*]. The time delay is a measure of the baseline vector: the straight line cutting through the earth from one telescope to the other. More precisely, the delay is proportional to the component of the baseline vector that lies along an axis from the earth to the quasar. To describe the vector completely in three dimensions, one need in principle only measure the time delays for three quasars whose directions in space establish three perpendicular axes.

In practice we typically make multiple observations of about 14 quasars from several telescopes. One reason for the practice is that we must allow for certain minute errors in the atomic clocks. When we begin an observing session, we do not know what the errors are; nor do we know precisely the positions of the quasars themselves. By making many observations and combining them through a least-squares adjustment (a mathematical procedure for finding the parameter values that best fit a set of data), we can obtain accurate estimates of these parameters and at the same time of the length and orientation of the baselines between observatories. The only limit on measuring a baseline is the requirement that the observatories at each end be able to "see" the same quasars simultaneously. In theory we could measure a baseline nearly as long as the earth's diameter (more than 12,000 kilometers), but in practice the greatest baseline distance we have recorded is about 10,000 kilometers.

We have now accumulated more than five years' worth of VLBI measurements that are accurate to within a few centimeters. It is important to realize, however, that five years is not a long time on the time scale of the geophysical phenomena we are studying. The results we shall discuss are of necessity preliminary ones, intended primarily to illustrate the great potential of VLBI.

This caveat is particularly apt in the case of a phenomenon that constitutes one of the oldest puzzles in geophysics: the wobble of the mantle and crust about the earth's rotation axis. As long ago as 1765 the Swiss mathematician Leonhard Euler showed that a rigid body would wobble in a periodic fashion about its rotation axis if the rotation axis were different from the body's axis of symmetry. In the following century it was calculated that the earth should wobble with a period of about 305 days. If it did, then the astronomical latitudes of all points on the earth, that is, their locations as measured from the positions of reference stars on the celestial sphere, should vary with the same period.

When the variation of latitude was finally detected in the 1890's by the American geodesist and astronomer Seth Carlo Chandler, it turned out to be more complex than had been predicted. It was found to have two components, one with a period of 365 days and an amplitude of about .1 second of arc (a distance of some three meters on the surface of the earth) and another with a period of roughly 430 days (14 months) and an amplitude of .15 second of arc. The 365-day term was easily explained as a forced motion associated with the seasonal redistribution of air and water masses. The 430-day component was assumed to be the free wobble predicted on the basis of Euler's work, but its period was embarrassingly different from the predicted value of 305 days. The problem, as the American astronomer Simon Newcomb soon explained, lay in the fact that the earth does not behave like a perfectly rigid body. When the elasticity of the mantle and the presence of oceans are taken into account, the calculated period of the wobble is found to be in reasonable agreement with the observed period of 430 days.

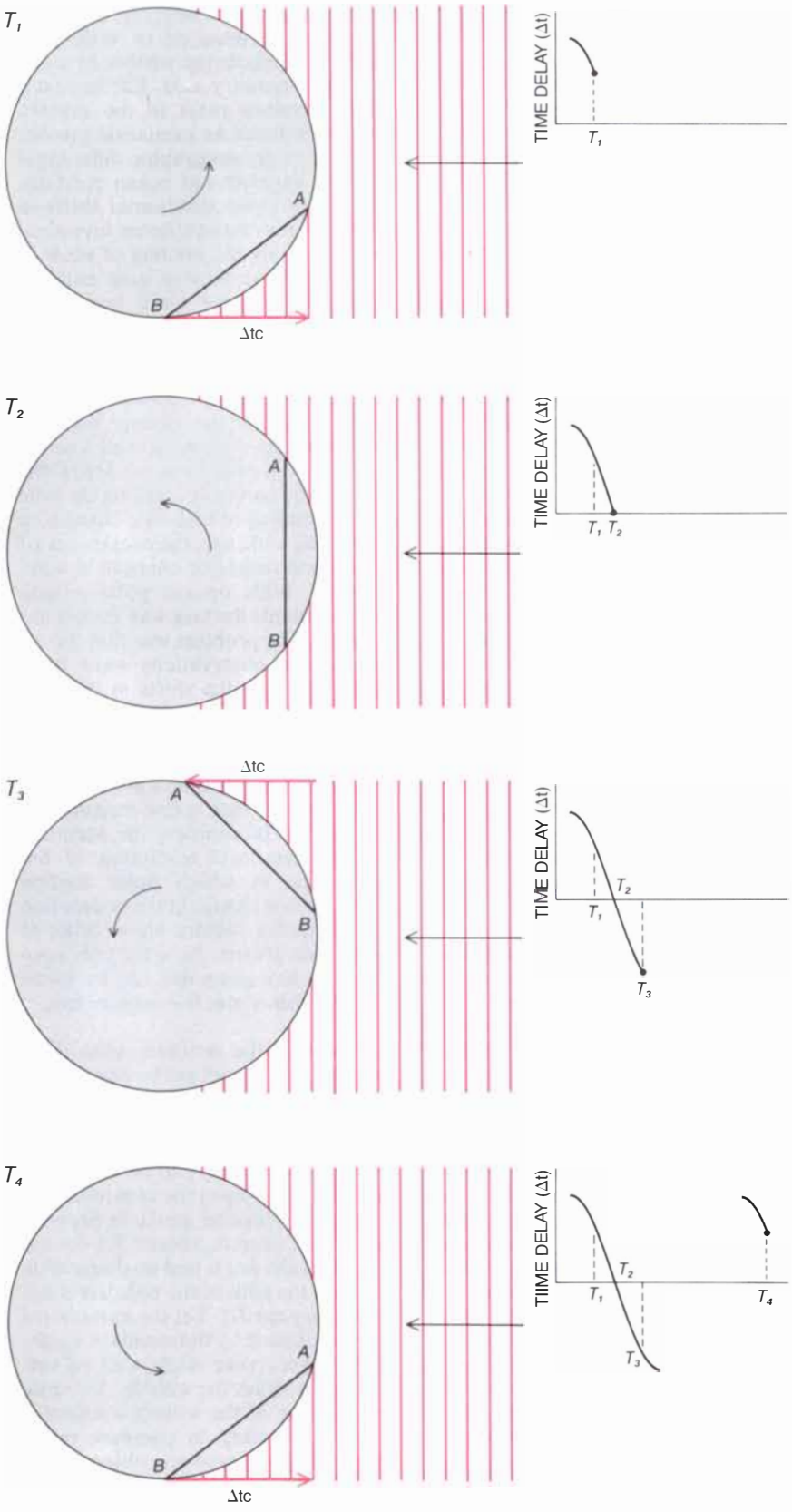
The 430-day wobble is now known as the Chandler wobble. Along with the forced annual wobble it is also referred to as "polar motion," because when the mantle and crust wobble about the spin axis, the pole traces a spiral path on the earth's surface. (Polar motion is not to be confused with nutation, which is a change in the orientation of the spin axis in space.) For the past 87 years a global network of observatories equipped with optical-mechanical instruments have tracked the wobble by measuring the changes it causes in astronomical latitude.

These measurements have shown the Chandler wobble is persistent, but they have failed to answer the fundamental question: What sustains it? The wobble is analogous to the oscillation of a pendulum. Just as the pendulum on a grandfather clock would slowly come to a stop if the energy it lost to friction were not replaced by energy from the lowering of the weights, so too would the Chandler wobble dissipate without a steady input of energy. The symmetry axis would gradually settle back into alignment with the spin axis, at which point the wobble would stop. Conversely, any large displacement of mass on or within the earth could excite the wobble by moving the symmetry axis. Earthquakes, which displace mass in the mantle, could constitute an excitation mechanism. So might geographic differences in wind patterns and ocean currents, which can cause substantial shifts in air and water masses. Some investigators also think the melting of snow in places such as Siberia may entail a mass displacement large enough to drive the wobble.

It is difficult to determine which of these mechanisms is the primary one because neither the amount of energy they release nor the amount required to sustain the wobble is well known. Yet one can still hope to settle the question by carefully tracking the pole and attempting to associate changes in its motion with, say, the occurrence of large earthquakes or changes in wind patterns. With optical polar-motion measurements the task was almost impossible. The problem was that the errors in the observations were large compared with the shifts in the pole expected from the various excitation mechanisms; one could never be sure whether an observed motion whose amplitude was less than a meter or that lasted for less than a few months was real or not. In contrast, the accuracy and the temporal resolution of the VLBI data, in which polar motion shows up as a change in the orientation of the baseline vectors, are an order of magnitude greater. Now the pole's position on any given day can be specified to within about five centimeters.

Even with the accurate VLBI data, however, it will not be easy to sort out the relative influence of the competing mechanisms for exciting the wobble. An individual earthquake, for example, would have to be very large indeed to show up in the VLBI measurements; the Mexican quake in September of last year registered 8.1 on the Richter scale, but it had no discernible effect on the path of the pole [*see illustration on page 52*]. Yet the cumulative energy released by thousands of earthquakes every year might well be sufficient to sustain the wobble. To settle the question of the wobble's source it will be necessary to compare many years of polar-motion observations with data on earthquakes, wind patterns and other possible mechanisms.

In the meantime the preliminary VLBI data contain one important result. Measurements made by optical observatories between 1899 and 1982 had suggested that in addition to the wobble, the pole displays a secular, or



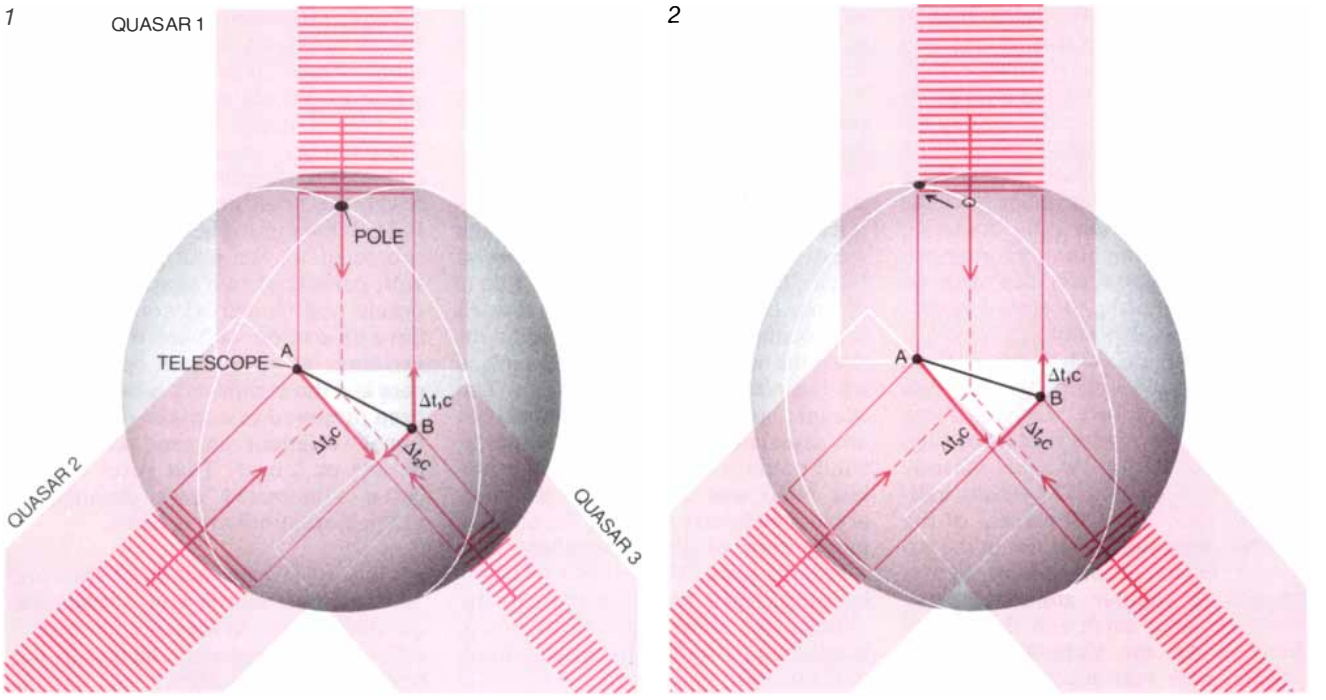
VLBI OBSERVATIONS detect the time delay between the arrival of a quasar signal at one radio telescope and its arrival at a second telescope. In this simplified illustration both telescopes are on the Equator. As the earth rotates, the time delay Δt follows a sinusoidal pattern. At time T_1 the signal arrives at telescope A first; at T_2 it arrives at both telescopes simultaneously, and at T_3 it arrives at telescope B first. At T_4 the earth has completed one rotation, and the time delay has returned to its value at T_1; the interval from T_1 to T_4 is exactly one day, and with VLBI it can be measured to within .1 millisecond. The amplitude of the time delay is proportional to the length of the baseline AB.

long-term, drift. Estimates of the rate and direction of the secular motion have varied widely; according to one estimate, the pole drifts about three milliseconds of arc per year in the direction of 70 degrees west longitude. Some workers, suspecting systematic errors in the optical measurements, have even doubted the reality of the secular motion. The first five years of VLBI data suggest the motion is real. The data yield an estimated drift of 3.7 milliseconds of arc per year in the direction of 45 degrees west longitude.

The importance of measuring the secular polar motion precisely has to do with its cause, which is thought to be a large, long-term displacement of mass in the mantle. Although the mantle responds to the sudden, comparatively weak stress of an earthquake as an elastic solid, it responds to stronger, long-term stress as a viscous fluid. One type of long-term stress arises from the advance and retreat of the glaciers that cover much of Canada and Scandinavia during ice ages. When the glaciers advance, they depress the crust, and large masses of rock in the underlying mantle are squeezed toward the Equator. During periods of glacial retreat the crust rebounds and mantle material flows back north. According to a widely accepted theory, the present northward flow causes the observed secular motion of the pole by displacing the earth's axis of symmetry.

The extent of the secular motion is determined by the rate of flow in the mantle, which in turn is determined by the viscosity of the mantle. Consequently the secular motion can be used to estimate mantle viscosity. The latter quantity is an extremely important one in geophysics: the plates that shape the earth's surface are thought to be propelled by a convective flow in the mantle. Mantle viscosity is notoriously difficult to estimate directly, however, because the composition, temperature and physical state of the mantle are not known well enough. One of the most widely accepted estimates is that of William R. Peltier and his colleagues at the University of Toronto, whose calculations are based on optical observations of the secular motion. When models such as Peltier's begin to incorporate the VLBI data, the accuracy with which the mantle viscosity is known should improve.

The same observatories set up early in this century to monitor polar motion revealed another phenomenon of geophysical importance: variations in the earth's rate of rotation. Historically the rotation rate had been assumed to be constant. Indeed, the solar day was the standard for measuring



LENGTH AND ORIENTATION of the baseline vector between two telescopes could in principle be specified in three dimensions by observing three quasars simultaneously. The time delay measured for each quasar is multiplied by the speed of the signal (the speed of light) to determine the component of the vector along the direction to the quasar (*1*). A movement of the earth's spin axis (nutation) or of its mantle and crust (polar motion) in relation to the stationary quasars changes the orientation of the baseline (*2*).

the passage of time, and clocks were simply mechanical devices for splitting the day into finer units. By the 1930's, as clocks were improved, it became clear that the rotation rate is not constant after all. The length of a day was found to vary annually; days in January are a few milliseconds longer than days in July.

In the 1940's and 1950's, as quartz-crystal clocks and later atomic clocks were introduced, other systematic variations in the length of day were noted, with periods of six months, one month and a fortnight. In part these variations were clearly tidal, that is, they were associated with changes in the gravitational attraction exerted by the sun and the moon on the earth. Because the orbits of the moon and the earth are well known, tidal variations in the length of day are readily predictable. The largest of them have amplitudes of a few tenths of a millisecond.

In addition to the tidal variations there are less predictable variations in the rotation rate, and it is in the study of these fluctuations that VLBI measurements are proving particularly interesting. Measuring the fluctuations is a straightforward task. As the earth rotates, the time delay between the arrival of a quasar signal at one observatory and its arrival at a second follows a sinusoidal pattern [*see illustration on opposite page*]. To determine the length of day one could simply measure the time delay at one moment and then measure how long the delay takes to return to precisely the same value. In practice we actually measure Universal Time (UT1), which is given by the rotational phase angle of the earth in space. An observing period of less than an hour with two radio telescopes suffices to determine the orientation of the earth in relation to the quasars and thereby to specify UT1 to within .1 millisecond. By measuring UT1 at approximately daily intervals and then subtracting successive measurements, we can compute the mean rate of rotation and the length of day.

The tidal variations in the rotation rate arise because the changing gravitational attractions of the sun and moon alter the shape of the earth and hence its moment of inertia. Since the total angular momentum of the earth (the product of its moment of inertia and its angular velocity) must remain constant, there must be a corresponding change in the planet's angular velocity. In contrast, the nontidal variations in the rotation rate are caused predominantly by exchanges of angular momentum among components of the earth. The mantle and crust—which rotate roughly in unison and whose angular velocity we measure with VLBI—can exchange angular momentum with the atmosphere, the oceans or the core of the earth. The question that has long interested workers in the field is which of these exchanges dominates the observed nontidal variations in the rotation rate.

As long ago as 1960 Walter H. Munk, now at the University of California at San Diego, and Gordon J. F. MacDonald, now at the Mitre Corporation, argued that the nontidal variations on a seasonal time scale were associated with changes in east-west wind patterns and therefore with a transfer of angular momentum between the crust and the atmosphere. More than a decade later Kurt Lambeck of the Australian National University and Anny Cazenave of the Centre d'Études Spatiales des Rayonnements in Toulouse, working with better weather data, showed this connection to be undeniable. Other nontidal variations, whose periods were either shorter than the seasonal variation (ranging from a few days to a few weeks) or longer (on the order of decades), remained to be explained. At the time most investigators thought the short-period variations were the result of systematic errors in the astronomical observations.

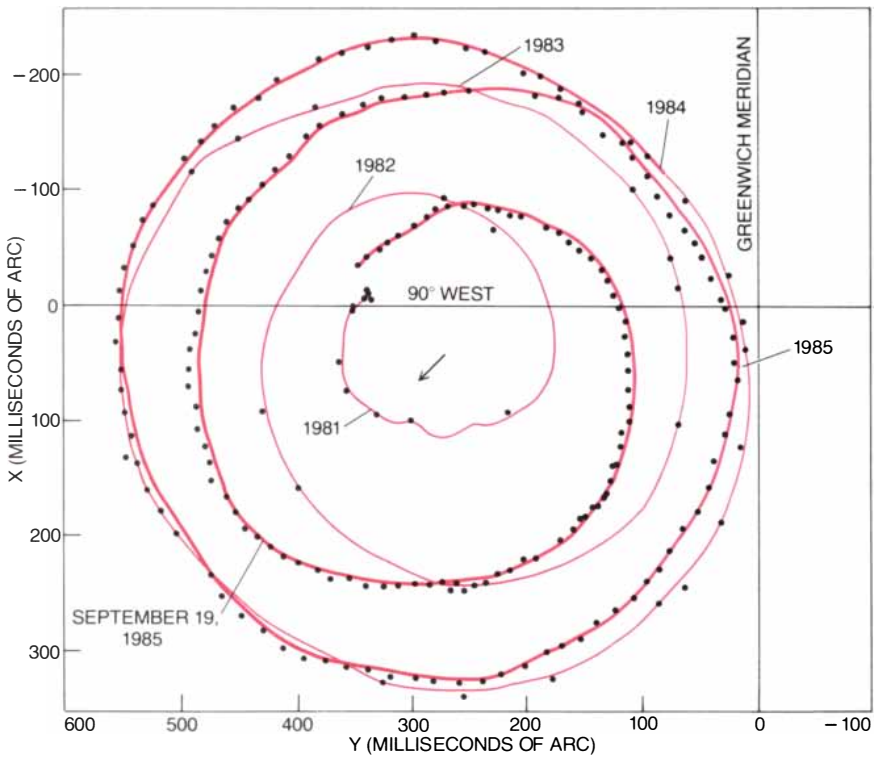
By analyzing global weather data with large computers and by employing VLBI measurements one can now conduct a more sophisticated analysis of the effects of the atmosphere on the earth's rotation. Richard D. Rosen and David A. Salstein of Atmospheric

and Environmental Research, Inc., in Cambridge, Mass., collaborating with workers at the National Meteorological Center, have developed a program that enables them to calculate day-to-day changes in the angular momentum of the entire atmosphere. From these changes one can infer changes in the solid earth's rotation rate, assuming total angular momentum is conserved. The inferred variations can then be compared with the observed variations in the length of day.

We have made such a comparison on the basis of the highly accurate length-of-day values derived from VLBI data during the period from July, 1981, through July, 1985 [*see illustration on opposite page*]. The results indicate that more than 90 percent of the seasonal, nontidal variations in the rotation rate are accounted for by exchanges of angular momentum between the solid earth and the atmosphere. When the globally averaged speed of east-west winds (mainly the subtropical jet streams) increases, the rotation rate decreases and the day gets longer. (The converse is of course also true.)

A dramatic example of this took place early in 1983, during the most recent occurrence of a climatological phenomenon known as El Niño. During El Niño, which recurs at irregular intervals of from two to six years, the westward trade winds in the equatorial Pacific fail; warm surface water that is ordinarily pushed by the winds toward the western Pacific sloshes back toward the east, warming the sea surface off the coast of South America. The oceanic and atmospheric anomalies are associated with global changes in wind patterns and weather. When the last Niño was at its peak in January and February of 1983, the angular momentum of the atmosphere also reached a peak. The transfer of angular momentum from the crust to the atmosphere slowed the earth down, lengthening the day by nearly three milliseconds.

The VLBI observations have also demonstrated that the nontidal rotational variations with periods shorter than a season, as short as a few days to a few weeks, are not merely errors in the optical measurements but are real. They too are caused primarily by exchanges of angular momentum between the crust and the atmosphere. Indeed, the atmosphere is the primary source of all the nontidal variations with periods of two years or less; the oceans and the earth's core appear to have little to do with any of them. An exchange of momentum between the core and the mantle has, however, long been proposed as a source of the rotational variations on time scales of a decade or longer. That issue will be better illuminated once decades of VLBI observations are in hand.



POLAR MOTION measured by VLBI is plotted for the period from September, 1980, to November, 1985. During 1980–81 the 365- and 430-day components of the wobble were out of phase and therefore tended to cancel. Hence the amplitude of the motion was relatively small. As the components moved into phase the amplitude increased, reaching a peak during 1983–84. The Mexican earthquake of September 19, 1985, had no discernible effect on the path of the pole. A complete wobble cycle, consisting of an outward spiral (*thin line*) and an inward spiral (*thick line*), takes about 6.5 to seven years. In addition to the wobble the pole displays a long-term linear drift. At the turn of the century the center of the wobble was at the origin of the plot, and it is now moving in the direction of about 45 degrees west longitude (*arrow*) at a rate of 3.7 milliseconds of arc per year. (A millisecond of arc corresponds to about three centimeters on the surface of the earth.)

Clues to the interaction of the core and the mantle can already be gleaned from VLBI measurements of a different phenomenon: changes in the orientation of the earth's spin axis in space. The changes are driven primarily by the action of the gravitational fields of the moon and sun on the earth's equatorial bulge. The plane of the Equator is tilted by an angle of 23.5 degrees with respect to the plane of the earth's orbit (the ecliptic), and whenever the moon and sun are away from the equatorial plane, they exert torques on the bulge that tend to pull it toward the ecliptic. As a result the spin axis precesses, describing a complete circle on the sky every 25,600 years. The angular radius of the circle is equal to the tilt of the spin axis, or 23.5 degrees. Superimposed on the precession are nutations, whose amplitudes are no more than nine seconds of arc—about 10,000 times smaller than the amplitude of the precession—and whose periods range from 4.7 days to one year to 18.6 years.

The gravitational torques that drive the nutations are well understood, but to predict the earth's response to those forces one must have a model of its internal structure. By comparing the predicted nutations with the observed nutations one can test the earth model. A model developed by John C. Wahr of the University of Colorado at Boulder was recently adopted as the standard for calculating nutations by the International Astronomical Union. The accuracy of Wahr's calculations exceeds that of the optical nutation measurements, and so his model can only be tested by VLBI measurements.

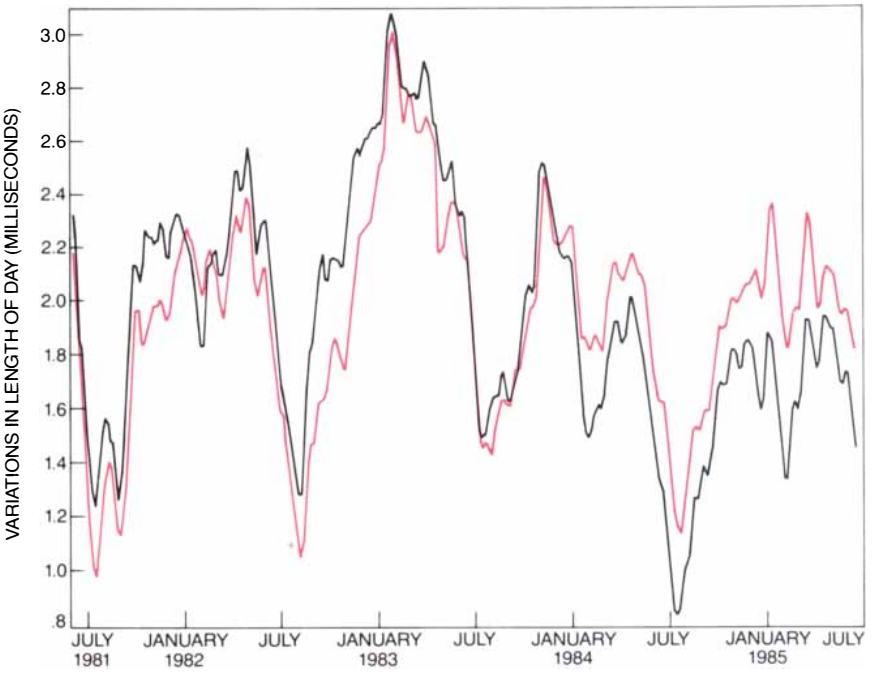
By analyzing VLBI data Thomas A. Herring, Carl R. Gwinn and Irwin Shapiro of the Harvard-Smithsonian Center for Astrophysics have detected an error of about two milliseconds of

arc in Wahr's value for the amplitude of the annual nutation. The error can best be explained, the workers argue, by supposing the earth's fluid outer core is slightly flatter than Wahr assumed it to be. (The core is ellipsoidal: its equatorial diameter is 6,971 kilometers but its polar diameter is only 6,954 kilometers.)

The explanation rests on the fact that the earth has a free nutation mode, one that is not forced by the gravitational fields of the sun and moon. Although this so-called free-core nutation has never been observed, its existence follows from the presence in the earth of an ellipsoidal, fluid core. Since the core is not rigidly bound to the mantle, its spin axis need not be in alignment with the axis of the cavity formed by the core-mantle boundary. If the two axes were to slip out of alignment, the core axis would precess about the cavity axis, and although the magnitude of the core's angular momentum vector would not change, the direction of the vector would. Because the total angular momentum of the earth must be conserved in magnitude and direction, the mantle would precess also, in the same direction as the core but 180 degrees out of phase. The motion of the mantle, which carries the crust with it, is in principle observable as a minute nutation of the earth's spin axis.

More important, the presence of the free-nutation mode affects the earth's response to the forces that drive the other nutations, including the annual one. The magnitude of the effect depends on the period of the free-core nutation, which Wahr's model predicts should be about 460 days. Herring and his colleagues have shown that a period 30 days closer to the period of the annual driving force would change the earth's response to the force by just enough to eliminate the two-millisecond error in the annual nutation. The period of the free-core nutation in turn is strongly dependent on the ellipticity of the core-mantle boundary. The Harvard group has calculated that a core only .5 kilometer flatter than the value assumed by Wahr would reduce the period from 460 to 430 days.

Such a subtle refinement is significant because the shape of the core is a critical parameter in models of the earth's interior; the anomalous flattening may be caused by the same convective processes in the mantle that drive the plates on the earth's surface. The refinement was made possible by the extraordinary precision of the VLBI observations, which monitor the nutations with an accuracy of better than one millisecond of arc. Considering how little is known about the core and about its interactions with the mantle, the VLBI measurements are a valuable new source of information.



LENGTH-OF-DAY variations measured by VLBI (*black*) correspond closely to the variations inferred from changes in the angular momentum of the atmosphere (*color*), demonstrating that the atmosphere has a significant influence on the earth's rate of rotation. The tidal variations in the rotation rate due to the gravitational effects of the moon and sun have been removed from the data. Exchanges of angular momentum between the atmosphere and the crust account for more than 90 percent of the nontidal, seasonal variations in the rotation rate; they also have a strong influence on shorter-period fluctuations. The large peak in the length of day in early 1983 coincided with El Niño. At that time angular momentum was transferred from the surface to the atmosphere; the earth slowed down and the east-west winds, notably the northern subtropical jet streams, intensified.

In contrast to the core, a wealth of information is available about the earth's surface and about the plates whose motions determine its topography. The relative motions of the plates are on the order of centimeters per year, and they are generally thought to be known to within a centimeter or less. Investigators such as Thomas H. Jordan of M.I.T. and J. Bernard Minster of Science Horizons, Inc., in Encinitas, Calif., have developed computer models that predict the rates of motion between points on different plates on the basis of evidence preserved in the geologic record. Yet the models make simplifying assumptions. They assume, for example, that the plates move at a constant rate and move rigidly, so that there is no motion within the plates. VLBI observations promise to put these assumptions to the test by directly measuring changes in the lengths of baselines.

The bulk of the observations collected so far have come from the POLARIS telescopes and collaborating observatories in Europe. The only plate boundary spanned by this network is the Mid-Atlantic Ridge, along which the North American and Eurasian plates are spreading apart while hot magma wells up from the mantle to form new oceanic crust. The relative motions along the baselines that cross the boundary are predicted to be less than two centimeters per year. VLBI measurements of the baselines between Westford, Mass., and telescopes in Onsala, Sweden, and Wettzell, West Germany, are consistent with the prediction: the baseline lengths are increasing at a rate of between one centimeter and two centimeters per year.

On the other hand, the baseline lengths also exhibit equally large random fluctuations; hence from these data alone we would be reluctant to conclude that we had really measured plate motions. Fortunately our confidence is buttressed by the parallel work of Robert J. Coates, Thomas Clark and their colleagues at NASA, who have recently begun a series of observations from a network of stations on the North American, Pacific and Eurasian plates. The motions

across the plate boundaries spanned by their network are for the most part expected to be substantially larger than the motion across the Mid-Atlantic Ridge. The measurements obtained by the NASA group are not significantly different from the predicted values. It appears that VLBI is indeed capturing plate tectonics in action.

Once we have accumulated enough observations to be sure the measurements of plate motions are accurate, the measurements will have great value as a check on plate-tectonic theory. Already one of our preliminary results seems to contradict an important assumption of the theory. The POLARIS observations indicate that the baseline between Westford and Fort Davis, Tex., is shrinking by about a centimeter per year. Theoretically there should be no change in the baseline length, since the two stations lie on the same supposedly rigid plate. There is still some doubt about whether the observed compressional motion is real; it could be the result of an as yet undetermined systematic error. And yet, as Steven A. Musman of the National Geodetic Survey and Thomas Schmitt of the Georgia Geological Survey have calculated, the strain produced by the compression of the crust would be consistent with the rate at which earthquakes occur in the eastern U.S. As VLBI observations continue, plate tectonics may turn out to be more complicated than many geophysicists had thought.

PLATE MOTIONS change the length of baselines between radio telescopes on different plates. The map of the Northern Hemisphere shows the baselines for which enough VLBI measurements have been accumulated to allow reliable estimates of the rates of change (in centimeters per year). The numbers in brackets are the plate motions predicted by a theoretical model. In general the predicted and measured rates are in rough agreement. An exception is the baseline between Fort Davis, Tex., and Westford, Mass.: theoretically there should be no motion between points on the same plate, but VLBI data indicate the baseline is shrinking.

WHY GMC TRUCK IS SAYING...

"IT'S NOT JUST A TRUCK ANYMORE."

The key word here is "just." "Just a truck" implies mere adequacy. Functionality and nothing more. That's the way trucks were. That's not the way GMC trucks are today. At all.

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

WE'RE CIVILIZED.

Time was when trucks had to hide out in the alleys and byways of the country making pickups and deliveries. You just couldn't take them anywhere.

That's not the case today. A GMC truck is bred for better things. It will still pridefully perform the work of our nation. But it will do so with a lot more style.

Light-duty GMC trucks come with the well-mannered niceties of a car. They offer air conditioning and stereo and plush seating and power this and power that. What's more, there's something about the lively "lemme at it" look of a truck that hints of an active and probably interesting lifestyle.

So, today, you will see GMC trucks perched proudly in the darndest places. Like the front of Restaurant Chez Posh. And the driveway of the rather impressive split-level at the end of your block.

GM

Trucks are the hottest-selling vehicles in the country today. And at GMC Truck, business has never been better. GMC trucks are preferred by tasteful buyers seeking quality and integrity.

WE'RE PROUD.

There was a time not so long ago, when a job putting together trucks was probably pretty unexciting. That's when trucks were just trucks.

Today, as the vehicles of preference for so many interesting buyers, trucks border on the glamorous. We can sense that feeling filtering from the showroom to the assembly line.

From the design engineers to the people who assemble to the final inspection teams, there's an infectious new sense of pride and purpose. Everyone knows they're producing something important: trucks that more and more people want today.

WE KEEP GOOD COMPANY.

GMC Truck today is a beehive of activity. We bring you heavy-duty Generals and medium-duty Top Kicks and buses for schools. We offer Jimmys and Safari Vans and small space-age pickups. All crafted with such enthusiasm, we're saying they're "not just trucks anymore."

Since we're not just trucks, we feel most comfortable associating with other vehicles that are not run-of-the-mill. So in addition to our GMC Truck dealers, our trucks are sold by selected Buick, Cadillac, Oldsmobile and Pontiac dealers. Four fine names indeed. Not "just" cars. Or... maybe their cars are sold by our fine GMC Truck dealers. Depends on how you look at it. However you choose to, just look in the Yellow Pages under "Truck." Or under "Automobile." We're listed under both. For increasingly obvious reasons.

GMC TRUCK

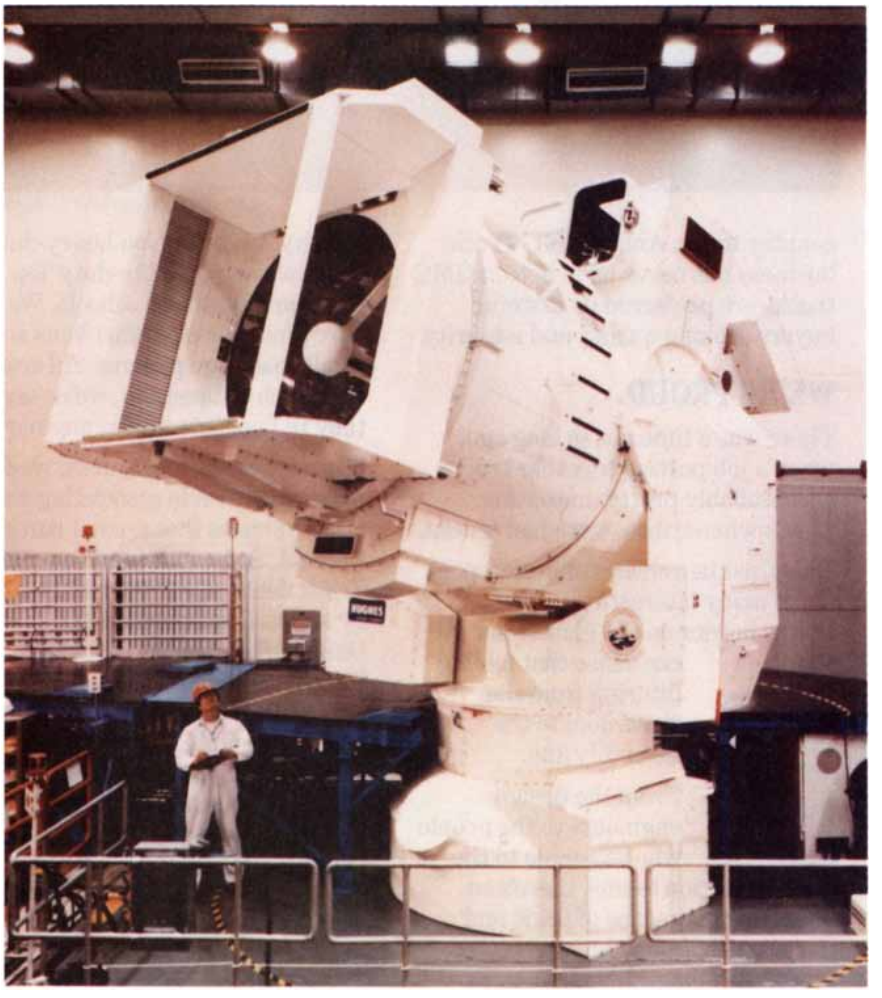
IT'S NOT JUST A TRUCK ANYMORE.

SCIENCE AND THE CITIZEN

SDI Boom – or Bust?

The Strategic Defense Initiative (SDI) is portrayed by the Reagan Administration as a research and development program designed simply to determine whether a defense against ballistic missiles is feasible. This modest mission definition belies the fact that the agency in the Department of Defense empowered to distribute funds for SDI-related R&D, the Strategic Defense Initiative Organization (SDIO), has already paid out $6 billion and has asked for the authority to disburse another $37 billion over the next five years. This makes the SDI the Government's fastest-growing R&D venture, and it could possibly go on to become the largest weapon-system program ever undertaken. Whether or not an effective defense system will come of it is open to question. Yet by virtue of its size (particularly if it advances to the production stage, which would raise the stakes perhaps as high as $1 trillion) the SDI is more than just another weapon system; it constitutes a de facto nationwide industrial program. Therefore a question equal in importance to that of the SDI's feasibility and reliability must be weighed: Is the SDI program in the best interests of the U.S. economy?

SDI supporters maintain that such a large-scale R&D program is precisely what the country needs to reinvigorate its industrial sector—that the massive investment in R&D by industry, Government laboratories and universities will guarantee the leading position of the U.S. in science and technology. The technological spin-offs from military R&D can be commercialized to generate enormous profits. A study done by the Business Communications Co., Inc., and touted by SDI advocates estimates that the total sales arising as by-products from SDI R&D could be as high as $20 trillion over the next 40 years. The study assumes, of course, that the SDI will proceed to full-scale production of a defense system.



LASER-BEAM DIRECTOR was developed by the Hughes Aircraft Company as part of the Strategic Defense Initiative (SDI) program. The director is designed to track objects in flight and to focus a high-power laser beam (generated by a separate device) on selected targets among them. Although the director was built for its military applications, SDI supporters argue that the technology it incorporates could have civilian applications.

The SDI program currently provides industrial companies, particularly the ones relying on the Department of Defense for a major share of their revenues, with substantial funds for R&D. (More than 75 percent of the SDI contracts awarded so far were distributed to just 20 of the top private defense contractors.) The greatest profits in any weapon program are made after the system has been approved for production. It is therefore understandable that defense-industry companies have contributed money to congressional candidates who might support the continued funding of the SDI.

A study done by the Council on Economic Priorities (CEP), an independent public-policy research organization, indicates that 19 leading SDI contractors, mostly aerospace and electronics companies, have contributed almost $6 million to various political-action committees (PAC's), which can raise unlimited funds on behalf of candidates for Federal office. The perfectly legal PAC contributions from defense-industry companies, the CEP maintains, are directed to "PACmen": candidates who hold positions of influence on congressional committees dealing with defense issues or whose constituencies depend on military contractors for jobs.

Although the SDI may be good for the top dozen or so defense contractors, many observers think it may not be quite so good for the U.S. economy as a whole. Edward E. David, Jr., former president of the Exxon Research and Engineering Company, points out the deleterious effects a "megaproject" such as the SDI can have on the nation's R&D. The "technological pizzazz" of such huge projects, David says, tends to attract scientists and engineers from smaller-scale R&D enterprises in such private-sector activities as the manufacture of capital goods and commercial products. Unfortunately, David maintains, it is precisely this smaller-scale, more mundane R&D that American industry needs in order to improve its economic competitiveness in world markets.

According to David, weapon-related R&D is concerned with the perfor-

mance of a technology rather than with its economical large-scale production. "Research for SDI," he says, "puts principal emphasis on performance, because the eventual SDI products will be manufactured on a much smaller scale than consumer items." David continues: "If the goal is to improve the manufacturing capability of the country, there certainly are more efficient, directed ways to do this than to rely on the 'fallout' from the SDI program."

Quick commercial application of SDI R&D could well be hampered by the classification of many of its results. Andrew M. Sessler of the Lawrence Berkeley Laboratory, Donald Prosnitz of the Lawrence Livermore Laboratory and their colleaques had been working on a free-electron laser, supported by the U.S. Department of Energy as part of the fusion-energy program. When the principal investigators accepted supplementary funding from the SDIO, enabling their project to proceed much faster, there was no mention of restrictions on publication of the group's work. It therefore came as a surprise when the work, directed toward developing a new way to increase laser power output sharply, was classified last year by the SDIO.

Although the SDIO did change its mind about the work of the group after a year and has since tried to present itself as being accommodating, SDI contractors are nonetheless subject to restrictions at the discretion of the Defense Department. In this context Robert N. Noyce, vice-chairman of the Intel Corporation, notes that the Pentagon often has taken over promising fundamental technologies for its exclusive use. The very-high-speed integrated-circuit program, Noyce says, "was originally justified on the basis that it would have commercial potential and benefit. But right now the Defense Department is trying to classify the whole thing."

Another factor that works to the detriment of spin-off commercialization is that the products the SDI might yield are not likely to be marketable, at least in their first incarnation as weapon-system components. Nearly half of the program justifications for SDI research submitted to Congress were prototype demonstrations of hardware for finding, tracking and destroying ballistic missiles. Such narrow, applications-oriented military research is less likely to have spin-offs than more basic research.

The narrowness of the scope of SDI R&D is also troubling universities that accept Pentagon funding for research. Carl Kaysen, director of the Program for Science, Technology and Society at the Massachusetts Institute of Technology, points out that traditionally the universities and the military have agreed that "the advancement of pure science and general technical expertise was a good thing.... Now the question is, are the security agencies going to narrow their focus?" Kaysen chaired a committee that conducted a survey of attitudes toward funding of defense research among the faculty at M.I.T., which has long been a prime recipient of such funding. The survey's results indicate that constriction of R&D to certain specific technologies has generated widespread unease among the M.I.T. faculty. Only 5 percent of the respondents said the SDI would be generally good for the university. Three-fourths of them thought the SDI would channel research away from the civilian economy, and 57 percent thought the SDI would not have a good effect on the economy.

Yet the Defense Department is fostering dependence on the SDI in academia too. According to the Council on Economic Priorities, Pentagon funding at universities has increased 115 percent over the past six years. As the largest and fastest-growing defense research program, the SDI will accelerate that trend. Soon the SDI may be the only funding option available for certain university researchers. Increasing the involvement of universities, along with defense contractors, in the SDI program may serve to legitimize and promote the program regardless of its technical merits or ultimate economic consequences.

Who Lost China?

Is China's family-planning program coercive? The evidence is scanty, scholars disagree and there are signs that the program has relaxed its stringency in the past three years. Nevertheless, the U.S. Agency for International Development—under pressure from conservative members of Congress—has decided to withhold the U.S. contribution to the United Nations Fund for Population Activities on the grounds that the UN organization supports the Chinese birth-control program and that the program is coercive.

The U.S. agency, generally referred to as AID, was to have given the Fund for Population Activities (UNFPA) $25 million in 1986, or about 20 percent of the fund's total budget. The money was withheld because AID officials interpreted legislation written by Representative Jack Kemp of New York and Senator Daniel K. Inouye of Hawaii as proscribing U.S. contributions to the UN group. An amendment added by Kemp and Inouye to a foreign-aid appropriations bill prohibits U.S. assistance to any organization that "participates in the management" of a program involving involuntary abortion or coercion.

Rafael M. Salas, executive director of UNFPA, contends that "the reasoning on which this [the AID decision] is based is factually wrong." According to Salas, "countries do not allow us to manage their programs." Little of the roughly $10 million that UNFPA gives to China is applied directly to family-planning programs, Salas said. The majority of the money is spent on computers for analyzing the results of the recent Chinese census, on scholarships that enable Chinese scientists to visit the U.S. to study demography and population policy and on communications and education within China pertaining to population matters.

AID administrators acknowledge that UNFPA does not provide family-planning services. Nevertheless, in 1985, the year the Kemp-Inouye amendment was enacted, it was determined by AID that the UN agency was "too close" to the Chinese program, according to an AID official who asked not to be identified. "We determined," the official said, that "UNFPA was participating in management, because they were providing the census work and training people for the family-planning program." As a result $10 million was withheld from the U.S. contribution to UNFPA last year. This year, he said, AID decided that neither the Chinese program nor the role of UNFPA had changed, and so the entire U.S. contribution was withheld.

Yet at least one U.S. China scholar believes Chinese population policy has become more lenient since 1982–83, when, at the height of the "One-Child Family" campaign, there were reports of forced abortions and sterilizations and even female infanticide. According to Susan Greenhalgh of the Population Council, Chinese population policy began moderating in 1984, partly because of popular resistance. In a recent issue of *Population Notes,* published by the Population Council, Greenhalgh reports that the circumstances in which Chinese couples are allowed to have more than one child have been broadened. "Crash" family-planning programs have been replaced by more routine efforts.

According to Greenhalgh, particular attention has been given to reducing the number of abortions, and the effort appears to have had some success: a limited survey of Chinese localities showed a decrease in the number of abortions between 1984 and 1985. "The situation has definitely been

liberalized and relaxed," Greenhalgh said recently. But she stressed that her data, like all data on Chinese population policy, are sketchy at best. "Nobody can say there is no coercion in China," she added, "because nobody has the data."

MEDICINE

AIDS: The Good News

At last there is some good news about AIDS. Although there is so far no prospect of a cure, two other goals seem somewhat less remote: a palliative drug, which might slow the course of the disease in some of its thousands of victims, and a vaccine, which could protect the hundreds of thousands who are at risk of infection.

Azidothymidine (AZT) is the first of many candidate antiviral drugs to have shown some therapeutic promise in AIDS. In September the manufacturer, the Burroughs Wellcome Co., called an early halt to a clinical trial of the drug in patients who had recently suffered their first bout of *Pneumocystis carinii* pneumonia (a common infection among AIDS victims) because an independent review board concluded it would be unethical to continue withholding AZT from the control group. Of the 145 patients given AZT, only one died in the course of the trial, which began in February. The treated patients also developed fewer infections and cancers than the 137 patients who were receiving a placebo, among whom there were 16 deaths.

AZT is believed to inhibit a crucial step in the life cycle of the AIDS virus, which is a retrovirus. The genetic material of retroviruses is RNA; in an infected cell it is "reverse-transcribed" into DNA, which then directs the production of viral proteins. AZT is thought to block reverse transcription by acting as a "fraudulent nucleoside": its close resemblance to deoxythymidine, one of the subunits of DNA, allows it to be incorporated into the growing DNA chain, but a crucial chemical difference then prevents the next link from being added. Reverse transcription fails, and the virus cannot replicate itself.

It is thought the drug's ability to inhibit DNA polymerization also accounts for the most serious of its side effects: the suppression of blood-cell production in the bone marrow. That is a drawback AZT shares with many drugs administered in cancer chemotherapy—the role for which it was developed in the early 1960's by Jerome P. Horwitz of the Michigan Cancer Foundation. Nevertheless, the clinical results have led the National Institutes of Health and Burroughs Wellcome to make the drug available to a broad group of AIDS patients.

The need for a palliative AIDS drug might one day be obviated by the development of a vaccine. That quest may have advanced a step with work described in *Nature* by a group of investigators led by Joyce M. Zarling of Oncogen in Seattle, who studied the antigenic properties of an experimental virus developed earlier by several members of the group. By means of recombinant-DNA techniques the workers had inserted the gene encoding the AIDS virus's protein envelope into vaccinia virus, the virus that serves as smallpox vaccine. The hope was that cells infected with the recombinant vaccinia virus would express the envelope proteins, which in turn would stimulate an immune response to the AIDS virus. A recombinant vaccine would be much less risky than one composed of the AIDS virus itself, even in killed or weakened form. Cells infected with the recombinant vaccinia virus did express the added gene as protein, and mice immunized with the virus made antibodies to the envelope proteins.

The Seattle group now reports that the recombinant virus stimulates not just antibody production but the full range of immune responses; moreover, it does so in primates. Like the mice of the earlier study, macaque monkeys immunized with the virus made antibodies able to react with AIDS-virus proteins. Antibodies are the product of the lymphocytes, or white blood cells, known as *B* cells. The workers were able to show that the monkeys' *T* cells, the other class of lymphocytes, were also stimulated by the envelope proteins encoded by the recombinant virus: they multiplied rapidly when they were later exposed to the AIDS virus, whereas *T* cells from unimmunized monkeys did not.

T cells have varied duties. Some of them orchestrate the immune response by secreting a substance called interleukin-2, which causes *B* and *T* cells to mature and proliferate. Others, known as cytotoxic *T* cells, actively destroy infected cells. The investigators have found that the recombinant virus stimulates *T* cells to carry out both tasks. Cells from macaques immunized with the recombinant organism secreted interleukin-2 when they were exposed to AIDS virus, but cells from control animals showed no such response. In a separate study, done with Jorg W. Eichberg of the Southwest Foundation for Biomedical Research, chimpanzees were inoculated with the recombinant virus. In the apes' blood the workers later found cytotoxic *T* cells that responded to the AIDS proteins.

That finding may have special relevance to the prospects for a vaccine. Human beings exposed to the AIDS virus produce antibodies that can react with the viral envelope, and yet the antibodies seem to have little protective effect. There is reason to think *T*-cell immunity might be more effective at warding off infection, according to Zarling. The AIDS virus causes infected cells to fuse with uninfected ones. The virus can thereby pass from cell to cell without being released into the serum and exposed to circulating antibodies. Conceivably *T* cells primed by vaccination might be able to attack the virus in its cellular refuge.

Cold Comfort

One of the commonest viral infections is also one of the most intractable: the common cold. The body develops no immunity to colds, and there is little hope for a cold vaccine, which would artificially stimulate immunity. The immune system cannot ward off colds because their agents are so diverse: most colds are caused by rhinoviruses, but that group includes at least 100 strains. The exposed regions of the rhinovirus capsid, or protein shell, to which antibodies bind are precisely the regions that vary the most among the strains. Antibody developed in response to one rhinovirus strain is of no use against other strains.

A defense against the full range of rhinoviruses might nonetheless succeed by exploiting hidden similarities among the strains. In 1984 Richard J. Colonno and Gordon Abraham of Merck Sharp & Dohme reported that the majority of the rhinovirus strains they tested bind to the same kind of cell-surface molecule when they infect human cells. The finding suggested a broad defense against rhinoviruses could take advantage of the common cell-surface receptor. Another possibility emerged last year when Michael G. Rossmann of Purdue University and his colleagues described the rhinovirus structure in atomic detail. The description, based on X-ray crystallography, showed that certain protein sequences common to all rhinovirus strains lie at the base of a deep canyon scoring each face of the capsid. The narrow opening would exclude the large molecules of antibody, but smaller molecules of an antiviral drug might reach the common sequences. Both possibilities have now been borne out.

In work described in *Journal of Virology* and at this year's meeting of the American Society of Microbiology,

Colonno and his colleagues devised a strategy for blocking the attachment of rhinoviruses to their major cell-surface receptor. The workers injected mice with human cell membranes, stimulating the animals to make antibodies against a range of cell-surface proteins. At least one kind of antibody was likely to be specific for the major rhinovirus receptor; added to a culture of human cells, the antibody would occupy the receptor sites and thereby protect the cells from a subsequent challenge with virus.

To find the antibody the investigators isolated and cultured antibody-secreting cells from the mice and tested each cell's product for the ability to block rhinovirus infection. Two of the 7,356 cell cultures tested yielded receptor-blocking antibody. In vitro the antibody proved to be extraordinarily effective at thwarting the virus.

The principle of receptor blockade has now been proved in vivo. In a small group of chimpanzees (the only animals other than human beings that can be infected with cold virus) pretreatment with antibody in the form of a nasal spray inhibited viral growth in the animals. A study of human volunteers, done in collaboration with Jack M. Gwaltney, Jr., and Frederick G. Hayden of the University of Virginia School of Medicine, also yielded encouraging results. None of the volunteers who had been given antibody was fully protected against infection, but their symptoms took a day or two longer to appear after exposure to rhinovirus than those of the control group and were only half as severe.

Colonno's receptor-blockade strategy would prevent rhinovirus from completing the first step in infection: binding to the cell surface. Ordinarily the virus then penetrates the cell and sheds its protein coat, baring the RNA that directs production of new virus particles. Current work described by Rossmann's group in *Science* points the way to the design of drugs that could stem infection by locking the rhinovirus particle in its coat.

A class of experimental compounds is already known to block the uncoating of rhinovirus and related viruses in laboratory tests. To determine how the compounds work the Purdue group studied the interaction of a rhinovirus and two of the substances, known simply as WIN 51711 and WIN 52084. The workers first crystallized the virus, a necessary preliminary to structural studies by X-ray diffraction, and then soaked some of the crystals in the antiviral compounds.

A comparison of diffraction data for the treated crystals and for control crystals showed that individual molecules of the WIN compounds find a binding site among the constant protein sequences forming the floor of each canyon. A channel leads from the canyon floor into the RNA core of the virus, and the WIN compound lodges partway down the passage, with much of the molecule nestled in a deep pocket in the channel wall and the rest protruding into the main channel and blocking it.

The result, the first atom-by-atom portrait of the interaction of an antiviral compound with a virus, suggests two ways of accounting for the WIN compounds' ability to prevent viral uncoating. In several plant viruses hydrogen ions must stream into the virus before it sheds its coat; if ion flow is also necessary for rhinovirus uncoating, the fact that bound WIN compound seals openings in the capsid might explain its stabilizing effect. Alternatively, the collapse of the pocket that accommodates the WIN molecule may be a crucial step in uncoating. By filling the pocket the WIN molecule may in effect brace the capsid and prevent its disassembly.

The result also suggests a pattern for new drugs that would interact with rhinoviruses in the same way as the WIN compounds do. Knowing the size and shape of the binding site and the nature of its interactions with the sample antiviral compounds, one could tailor drugs that would fit the niche as well or even better. Counterparts to the WIN binding site are known or suspected to occur on rhinovirus relatives, which include poliovirus and hepatitis-*A* virus. Rossmann and his colleagues raise the possibility of rational drug design against not only the cold virus but also its more virulent kin.

Seeing the Light

The biological clock that controls rhythms of thought, behavior and physiology in humans has (in most people) a period longer than 24 hours; only because the clock is reset each day by environmental stimuli do people wake in the morning and sleep at night. What resets the clock? In a case study reported in *Science,* Charles A. Czeisler and his colleagues at Brigham and Women's Hospital in Boston tell how timed exposure to light set forward a 66-year-old woman's clock by six hours—a resetting dramatic and rapid enough to suggest that light is the synchronizer of the clock.

Until recently it was assumed that the sleep-wake cycle governs the human circadian, or daily, rhythm; until 1980 there was no evidence that human cyclic behavior is at all influenced by light. Then studies began to show that exposure to a light stimulus at night affects the neuroendocrine system. It was still not clear which was the more powerful influence, the light itself or its interruption of the sleep-wake cycle. After all, people exposed to light do tend to wake up.

Czeisler's team found a way around this chicken-or-egg problem when they identified a 66-year-old woman whose unusual circadian period was actually shorter than 24 hours. They hypothesized that older people, whose internal clocks tend to run fast, might be particularly sensitive to light at dusk (while they are still awake) and that younger people, who have long circadian periods, are more sensitive late at night. Dusk light should then serve to reset fast internal clocks, lengthening them (setting the circadian period forward) to fit a 24-hour day, whereas dawn light resets slow clocks by shortening their period.

The investigators had their elderly subject follow her normal midnight to 6:00 A.M. sleeping pattern in a controlled environment. Then, for one week, she spent four hours (from 8:00 P.M. to midnight) sitting in front of a bank of lamps with an intensity 48 times greater than normal indoor light, an intensity comparable to sunlight at dusk. The artificial dusk signal at this late hour tricked the subject's internal clock (as measured by her temperature and the secretion of particular hormones) to reset six hours forward, from midnight to 6:00 A.M.

This suggests it is light—in ordinary circumstances sunlight—that resets the biological clock. If that is the case, Czeisler says, therapeutic doses of bright light may turn out to ease certain problems associated with jet lag, changes in shift work and other sleep-wake disorders. But what does a single case study prove? Well, Czeisler says, seeing a single elephant fly would make people think twice about accepted aerodynamic theory. Likewise a single case study showing such an unexpectedly large and rapid resetting of an internal clock must mean that light is a factor in the regulation of human circadian rhythms.

SOCIAL SCIENCE

Untrashing Margaret Mead

Three years ago the public reputation of Margaret Mead went into at least partial eclipse, the penumbra of which chilled the reputation of U.S. anthropology as well. In *The American Scholar* Roy A. Rappaport of the University of Michigan has now reviewed the event in a discussion that may be-

gin to restore both Mead and her science to their place in the sun.

The main cause of the decline of Mead's public reputation (and of a reexamination of it among her colleagues) was *Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth,* by Derek Freeman of the Australian National University. Freeman charged that Mead's famous *Coming of Age in Samoa* gave a distorted picture of the behavioral norms governing sexuality in Samoan society in order to support the view that cultural rather than biological factors are the supreme determinants of human behavior. Mead, he said, found that the repression, fear and guilt she considered to be endemic in industrial society were not reflected in Samoan attitudes or behavior, and she argued that attitudes toward sexuality and other profound human emotions must therefore be shaped by culture and circumstance, not by biology or genetic endowment. Freeman insisted that her conclusions, and the findings on which she based them, were wrong.

Freeman's attack found a powerful amplifier in the mass media. Even before the book was published, Freeman's arguments made the front page of the *New York Times.* Other newspapers throughout the country also ran the story. One editorial asked how anthropologists could claim to be scientists if they fitted their observations to the theoretical framework and value system they brought into the field along with notebooks, tape recorders and cameras. Ultimately Freeman's analysis found its way into magazines and electronic media.

In attacking Mead, Freeman focused on two main points. First, Freeman judged her fieldwork to have been incompetent: she spent too little time in the field, her knowledge of the Samoan language was inadequate, she chose not to live in a Samoan household and she was duped by her female adolescent informants. Freeman held that Mead had been inadequately prepared for her first fieldwork by her distinguished mentors, Franz Boas and Ruth Benedict. More important, Freeman maintained that Mead's guiding theory, which he labeled "The Boasian Paradigm," was a doctrine of extreme cultural determinism, a position that left no adequate room for biological factors in explaining human behavior.

Rappaport counters both arguments vigorously. Admitting that anthropologists now generally agree Mead's account of Samoa was too simple and one-dimensional, he nonetheless maintains that if she did not see the whole truth, she did not get it all wrong either. Freeman's work, on the other

Our Customers are the World's Smartest Buyers of Computer Products because they will SAVE $6 MILLION this year. . . Please join them . . . at ELEK-TEK

DATA STORAGE

i·MEGA

BERNOULLI BOX

For PC/XT/AT and most compatibles

10 meg (single) $1250
20 meg (single) 1599
20 meg (10 + 10) 1850
40 meg (20 + 20) 2399
Interfaces priced separately

IOMEGA Bernoulli Cartridge

10 meg—1 pak 52
10 meg—3 pak 149
20 meg—1 pak 65
20 meg—3 pak 185
BB head cleaning kit 69

Dysan—Bernoulli Cartridges

10 meg—1 pak 44
10 meg—3 pak 125
20 meg—1 pak 59
20 meg—3 pak 165

HARD DRIVE CARDS

Plus Development . . . . . CALL
Western Digital . . . . . CALL
Mountain . . . . . CALL

MULTI FUNCTION BOARDS

AST

Six pack + with 64K $159
Six pack + with 384K 225
Rampage PC 310
Rampage AT 465
Ram Vantage 215

GENERIC

Multi-function Board
(clock & calendar, 2 ser., 1 par.)
0K 110
384K 175

INTEL ABOVEBOARD

All products . . . . . CALL

ORCHID TECHNOLOGY

All products . . . . . CALL
Generic Compatibles of most of the above . . . . . CALL

CHIPS

64K ram (150 or 200 NS) $1.35
128K ram 3.00
Samsung 256K ram 2.60
Major Brand 256K 3.30
NEC V-20 or V-30 . . . . . 12.00
Intel 8087 math chip . . . . . CALL
Intel 80287 math chip . . . . . CALL

HARD DRIVES

SEAGATE complete kit

ST 225—20 meg for PC/XT . . . $450
ST 238RLL—20/30 meg for PC/XT 525
ST—30 meg for AT (drive only) 720
ST—40 meg for AT (drive only) . . CALL
All above are pre-tested

TAPE BACKUPS

Western Digital 60 meg . . . . . CALL
Archive 60 meg . . . . . CALL
Tallgrass 20 meg . . . . . CALL

MODEMS

Hayes

Model 1200

HAYES *Internal*

(including software)
1200B 1200 Baud $359
2400B 2400 Baud 525

HAYES *External*

(no software included)
1200 1200 Baud 380
2400 2400 Baud 599

OTHER MODEMS

Internal *(including software)*

Practical Peripherals 1200B 125
Novation 1200B 130

External *(no software included)*

U.S. Robotics 1200 baud 199
U.S. Robotics 2400 baud 399
Novation 1200 baud 145
Novation 2400 baud . . . . . CALL

HEWLETT PACKARD

TONER CARTRIDGE
Mfr. Sugg. Ret. $115
Elek-Tek Price $84

$360
FX 85

EPSON

Epson's FX-85 is the industry's standard DOT MATRIX PRINTER. Features include 8K buffer, NLQ, 60/160 cps.

OTHER EPSON PRINTERS

FX 85 160 CPS $360
FX 286 wide carriage 160 CPS . . 470
LX 86 120 CPS 230
EX 800 300 CPS 475
EX 1000 300 CPS 690
LQ 800 NLQ (Parallel) 499
LQ 1000 NLQ (Parallel) 699
LQ 2500 NLQ 324 CPS 1050
CR 420 NLQ 420 CPS 900

PRINT BUFFERS

PRACTICAL PERIPHERALS

64K par. or ser. $129
256K par. or ser. 159

HANZON

64K universal 250
64K memory expansion 56

QUADRAM

Microfazers . . . . . CALL

MONITORS

SAKATA

SG 1000 green composite $95
SA 1000 amber composite 105

ZENITH

ZVM124 (TTL) amber 105

AMDEK

310A (TTL) Amber 150
600C RGB color 390
722C 488
725C (for ATT) 560

PRINCETON

HX9 9″ RGB 450
HX12 12″ RGB 465
HX12E 12″ RGB 550
MAX12E 12″ amber 175

Huge Discounts on

TOSHIBA NEC
OKIDATA XEROX / Diablo

SPECIAL PURCHASE!
Diablo D-36 Daisywheel 35 CPS
Mfr. Sugg. Ret. $1495
Elek-Tek Price $450

POWER PROTECTION

Surge protector strips

Perma Power 6 outlet $30
Isobar 4-6 4 outlet 42
Isobar 8-15 8 outlet 50

Line conditioners/stabilizers

TRIPPLITE

TL 1200 4 out 1200W 135
TL 1800 6 out 1800W 199

Command consoles

TL cc 1812 command console . . . 95

Stanby power systems

TRIPPLITE

TLBC425 425 watt system 360
TLBC1000 1000 watt system 850

SAFE

Safe 400VA 400 volt system 475
Safe 1000VA 1000 volt system . . . 875

DATA SWITCH BOXES

SES25 25 pin A-B 2 pos. $49
SES36 36 pin A-B 2 pos. 54
SES-3p25 25 pin ABC 3 pos. . . . 54
SES-3p36 36 pin ABC 3 pos. . . . 60

VIDEO BOARDS

Hercules graphics PLUS . . . . . CALL
Hercules graphics $199
Generic compatible 90
Hercules color 159
Generic compatible 85
Quadram EGA 375
Generic compatible 245

HUGE SAVINGS ON DISKETTES *Call for Quantity pricing for 10 boxes or more*

(*Packed in flip n' file) (**unformatted)	Dysan	maxell	3M	ELEK-TEK	SONY	MEMOREX
3½″ SSDD	15.00	14.50	14.50*	—	15.00	—
3½″ DSDD	22.00	21.00	21.00	—	24.00	—
5¼″ SSDD	13.00	8.00	8.50	—	—	7.00
5¼″ DSDD	18.00	10.00	10.00	10.50	—	9.00
5¼″ SSDD96TPI	20.00	18.00	17.00	—	—	—
5¼″ DSDD96TPI	22.00	22.00	22.00	—	—	—
5¼″ DSDDHD (for IBM AT)	24.00	24.00	23.00	17.00	—	—
8″ SSDD**	19.00	21.00	20.00	—	—	—
8″ DSDD**	23.00	24.00	23.00	—	—	—

3M DATA CARTRIDGES

3M	1	10	25
DC100	$13.50	$13.00	$12.00
DC300A	17.50	17.00	16.00
DC300XL	20.50	20.00	19.00
DC300XL/P	21.50	21.00	20.00
DC600A	23.00	22.50	21.50
DC1000	14.50	14.00	13.50
DC2000	19.50	19.00	18.50

ELEK-TEK 6557 N. Lincoln Ave., Chicago, Illinois 60645 (312) 677-7660 (312) 631-7800

CALL TOLL FREE 800-621-1269—EXCEPT Illinois, Alaska • CANADIAN TOLL FREE 800-458-9133

Corp. Accts. Invited. Min. Ord. $15.00. Visa or MasterCard by Mail or Phone. Mail Cashier's Check, Mon. Ord., Personal Check (2 wks. to clear) Add $4.00 1st item. (AK, HI, P.R., Canada add $10.00 first item) $1.00 ea. add'l shpg. & handl. Shipments to IL address add 7% tax. Prices subj. to change. WRITE for free catalog. RETURN POLICY: Defectives Only: Most products replaced within 30 days of purchase with identical merchandise only. Computer and large peripherals replaced only when defective on arrival (within 3 work days of delivery). Other problems covered by mfr. warranty. ALL ELEK-TEK MERCHANDISE IS BRAND NEW, FIRST QUALITY AND COMPLETE. Delivery subject to availability. **DUNS #09-718-0517**

hand, is "even more oversimplified than the account to which it is reacting.... It is no more than a heap of observations constituting half-truths." If Mead relied heavily on conversations with young women, Rappaport asserts, Freeman's easy access was to middle-aged male community leaders. "She was largely concerned with the actual sexual behavior of adolescent girls..., Freeman with an ideology of sex held publicly by elite adults."

Rappaport points out that in attacking Mead, Freeman was actually going after even bigger game. "If Freeman's proximate target was Mead's vision of Samoa or, possibly, the public's vision of Mead, his ultimate aim may well have been to discredit her mentor, Franz Boas, and, through him, American anthropology down to this day."

With what Rappaport describes as an apparent display of fine impartiality, Freeman invoked Charles B. Davenport, an eminent early 20th-century biologist, as the intellectual counterpoise to Boas. Against what Freeman regarded as Boas' cultural determinism, he cited Davenport's strong advocacy of genetic determinism. Yet Davenport, Rappaport writes, went beyond a neutral biological determinism: he warmly approved of the views of Madison Grant and other eugenicists who regarded the influx of Eastern European, Middle Eastern and Jewish immigrants as a major threat to what they regarded as American genetic integrity. Freeman recognized the racist tendencies of the eugenicists, Rappaport writes, but treated the tendencies as mere flaws in a scientific enterprise; actually the movement was "one of the most vicious expressions of racism in American history."

Rappaport says it is misleading to assert that Mead and Boas advocated a purely cultural reading of anthropologic data. Neither Boas nor his students expounded a doctrine that "'totally excluded biological variables.'" To suggest that they did is to fail to understand Boas' anthropology and the anthropology that followed him.

"American anthropology," Rappaport writes, "does not divorce culture from biology.... Moreover, 'a view of human evolution in which the genetic and exogenetic are distinct but interacting parts of a single system,' as Freeman puts it in his book, has long been part of American anthropology's received wisdom."

Felonious Employment

The U.S. prison population stands at more than 500,000. By the time males living in urban areas reach the age of 18, from 25 to 45 percent of them have been arrested for a nontraffic offense. Between 12 and 18 percent of these young men were arrested for such serious crimes as murder, rape, robbery, aggravated assault or auto theft. The total of such offenses exceeds 13 million.

The implications of these statistics for public safety and for the effectiveness and cost of the criminal-justice system prompted the National Institute of Justice, the research arm of the Department of Justice, to ask the National Research Council to commission a panel of 13 social scientists and three criminal-justice officials. Under the leadership of Alfred Blumstein of the School of Urban and Public Affairs at Carnegie-Mellon University, the panel looked for ways to reduce the frequency of criminal behavior and control the rapidly escalating costs of imprisonment.

The panel's report, *Criminal Careers and "Career Criminals,"* shows that certain social and environmental factors correlate most strongly with participation in crime. A criminal is likely to have received inconsistent or sporadically violent discipline from his or her parents; parent-child communication may have been poor and the relationship of the parents troubled. The criminal's family is likely to be large and one or both parents are likely to have been involved in crime.

Criminal activity also tends to be associated with (but is not necessarily preceded by) antisocial behavior and poor school performance. In the mid-teen years a frequent offender is likely to have friends who are delinquent; he or she is likely to abuse hard drugs, a pattern that persists into adulthood.

So far, so familiar. Surprises emerge from the study as it focuses on the factors that are linked to high or low frequency of offense as distinct from participation in crime. Blacks, for example, are at least twice as likely as whites to be arrested for at least one violent offense by the age of 18. Yet once an individual becomes an active offender the racial imbalance in the statistics largely disappears: active black offenders and active white offenders commit a similar number of crimes per year.

The same shift affects the statistics pertaining to male and female offend-

MARGARET MEAD is seen with children on Manus in the Admiralty Islands in 1928, the year in which her best-seller on Samoa was published. UPI/Bettmann Newsphotos.

ers. Men are from two to three times more likely than women to be arrested for a single offense, and yet the annual arrest rate for active female offenders is comparable to that of active male offenders. Such factors, the panel cautions, make dependence on demographic variables in order to target career criminals for selective sentencing ineffective as well as unethical.

Criminal careers (like those of professional athletes) tend to be relatively short: in various studies the average ranges from five to 15 years. Participation in crime typically peaks at about age 17 and then drops sharply. Conventional wisdom in the field therefore holds that the termination rate beyond the age of 30 must be high, because few active offenders are that old or older. Yet the panel found offenders who started young and remain active in their thirties "have the lowest termination rates and the longest remaining careers." Presumably such offenders have advanced beyond the reach of factors that end most criminal careers at earlier ages. The factors remain to be more clearly defined but seem to include marriage and legal employment, emotional and physical burnout, and death.

The panel concludes that a variety of measures might modestly reduce the crime rate. Certainty of sentencing for individuals who commit crimes frequently, who have an early juvenile-court record and who often abuse drugs could reduce indexes of serious crime by from 5 to 10 percent. Such a sentencing policy, the panel says, would produce a rise of 10 to 20 percent in the prison population, whereas a general increase in imprisonment would double the inmate population to achieve a comparable reduction in the crime rate. Some preventive measures, such as Head Start programs, may also be promising, according to the panel.

Homeless in Chicago

The ongoing debate about the ongoing tragedy of homelessness in America is only rarely enlightened by solid facts. A wealth of them are offered in a study done recently in Chicago, where investigators systematically combed shelters and streets in an attempt to get a fix on the shifting homeless population. The results confirm a bleak and familiar picture of isolation, institutionalization and unemployment (and unemployableness), as well as paralyzing poverty. Yet the study also offers grounds for optimism. In Chicago, at least, the homeless seem to be fewer in number, and so their problem seems to be potentially more manageable, than had previously been thought.

The study was carried out by Peter H. Rossi and his colleagues at the Social and Demographic Research Institute of the University of Massachusetts at Amherst, in collaboration with investigators at the National Opinion Research Center of the University of Chicago. It consisted of two surveys, the first one done in the early fall of 1985 and the second in February and March of this year. During both surveys workers from the center visited most of the shelters in the city, counted the names on the registers and interviewed selected residents. Meanwhile other workers searched a sample of city blocks, entering "each structure that allowed public access"—including abandoned buildings, bus terminals and all-night movie theaters—and attempting to interview "each person encountered in that search." According to Rossi and his colleagues, the combination of rigorous sampling methods, intensive searches and interviews to identify homeless people distinguishes the Chicago study from the few previous attempts to count the homeless.

What all such surveys have in common is that they tend to yield estimates of the homeless population that are much lower than the ones put forward by advocacy groups on the basis of less systematic research. In Chicago, it now seems, between 1,600 and 3,100 people are homeless on any given night. People drift in and out of homelessness, however: between 4,000 and 6,000 people experience bouts of it during the course of a year. Yet before the recent study was completed the number of homeless in Chicago had been put as high as 25,000.

Who are the homeless? In Chicago three-fourths of them are men; their average age (40) is about that of the general population. Although disproportionately few are college graduates, by and large they are not less educated than other people. More than half are black. A majority (57 percent) have never been married, and more than 90 percent are not currently married and are not accompanied by families. One-third have no contact whatever with relatives; nearly one-fourth have no contact with either relatives or friends. Nearly half of them seem to be severely depressed; nearly 15 percent show symptoms of psychosis, and some 23 percent have been hospitalized for psychiatric care.

From this general picture distinct subgroups emerge. A significant minority of the homeless are young women, often with children, who seek temporary refuge in shelters from "intolerable marriages or other living

arrangements." On the other hand, "homeless men under 40 [have] clearly been in trouble." More than three-fourths of them have been in mental hospitals or detoxification centers or have had some kind of criminal-justice experience (prison, jail or probation).

The common denominator is poverty. The homeless in Chicago tend to have been without steady work for much longer than they have been without homes—on the average, for more than four years. Their average income is $168 per month. Virtually all of them appear to be eligible for some form of government assistance, and yet fewer than 40 percent receive it.

Given the unexpectedly small number of homeless people in Chicago, it would seem possible at least to lighten their burden. General Assistance, the basic welfare program, pays a maximum of $154 per month. Rossi and his colleagues point out that the General Assistance rolls in Chicago now carry slightly more than 100,000 individuals, and they go on to remark that "an additional 1,000 to 1,500 recipients would scarcely swell the rolls to an unbearable size."

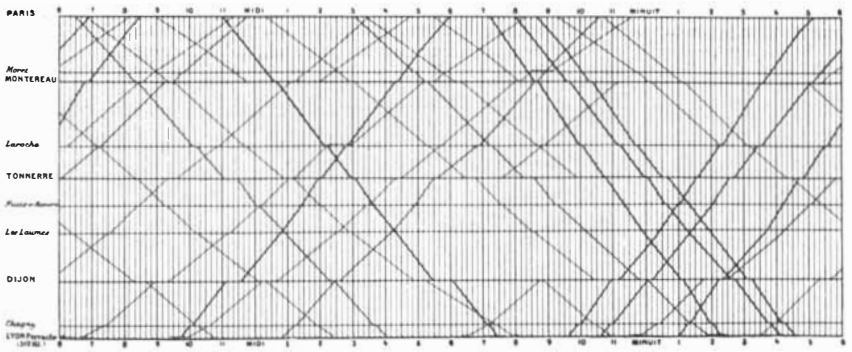
BIOLOGY

Can the Panda be Saved?

The giant panda, whose range is now confined to a few bamboo forests in western China, is close to extinction. The danger is so immediate that China has established a dozen reserves, supporting half of the species' estimated 1,000 individuals, and has trained emergency rescue squads. In collaboration with the World Wildlife Fund, China has also embarked on a major research effort to bolster the animal's breeding success and to develop other ways of reversing the species' decline.

The severest threat to the panda's future is territorial encroachment by human beings, which has forced the panda into isolated habitats and limited its supply of food and mates. At the same time a widespread dying out of bamboo, the panda's staple diet, has caused many pandas to starve.

At varying intervals averaging approximately 50 years different types of bamboo (which normally keep producing new shoots from a thick underground stem) flower, produce seeds and die. After such a die-off the affected bamboo species is not a viable food source until new plants mature in seven or more years. A die-off in 1983 involved at least two kinds of bamboo favored by the panda. In the past pandas could usually leave their mountainside territories and find alternative



This remarkable graphical train schedule, drawn in 1878 by the French scientist E. J. Marey, shows the trains running between Paris and Lyon, with intermediate stops. For example, the first train of the day leaves Paris at 6:30 am (upper left corner). The diagonals moving from upper left to lower right show all trains from Paris to Lyon. Trains returning from Lyon up to Paris are indicated by diagonals running from the lower left to the upper right. Steeper lines indicate faster trains. When a train stops at a station, the line is horizontal during the time of the stop. When two trains pass one another, the lines on the graphical train schedule cross. Graphical schedules are now used in Japan for managing the day-to-day operations of the *Shinkansen* (the bullet train) and by New Jersey Transit for public timetables.

bamboo patches at lower altitudes. Now much of that low-altitude land has been claimed by farmers and others, leaving many pandas with nowhere to go—and with no food.

In addition to severe habitat destruction and the flowering of bamboo, other factors contributing to the giant panda's endangerment include poaching and such reproductive problems as a high infant-mortality rate and a low birthrate; female pandas, which typically produce one or two cubs at a time, can be expected to raise only one cub to maturity in three years. Moreover, inbreeding may be common within the many habitats that support small (usually fewer than 40 individuals) populations. Inbreeding can promote the expression of deleterious genes, such as ones that make an animal subject to disease or that impair reproductive efficiency. It can also reduce the variety of genes in the population, a phenomenon that may hinder the ability of a species to adapt to environmental change.

In the collaborative research effort, initiated in 1980, investigators from China, the U.S. and elsewhere are studying the panda's reproductive cycle and mating behavior in order to help overcome the extreme difficulty the species has reproducing in captivity. If mating and artificial-insemination programs are successful, the cubs could be introduced into the wild to increase the size of the existing populations. Workers have also been evaluating the panda's requirements for survival in nature so that any animals born in captivity or rescued from inadequate habitats can be placed in optimal environments.

This past September China and the World Wildlife Fund approved an additional research project, which is to evaluate the genetic diversity of the species. They invited Stephen J. O'Brien of the National Institutes of Health and John A. Knight of the Doha Zoological Gardens in Qatar to collect blood samples, and possibly tissue samples, from pandas in nature reserves. From these samples the investigators plan to identify and compare the animals' alleles (gene variants) at some 60 genetic loci. Similar studies by O'Brien have revealed that the cheetah, which is also an endangered species, has almost no genetic variability and therefore may be highly inbred (see "The Cheetah in Genetic Peril," by Stephen J. O'Brien, David E. Wildt and Mitchell Bush; SCIENTIFIC AMERICAN, May). If analyses of panda samples suggest the panda too is seriously inbred, managers of China's nature reserves may decide to apply new strategies to maximize outbreeding among the various panda populations. "Mankind," notes O'Brien, "has learned a great deal from the study of animals. Now it is appropriate for the flow of knowledge to reverse direction to help an endangered species."

GIANT PANDA in the U.S. National Zoological Park feeds on bamboo, the staple diet of the species. The panda, whose range is now restricted to bamboo forests in western China, is seriously endangered, largely because of encroachment on its habitat by human beings. A widespread die-off of bamboo has added to the pressure, as has poaching. The photograph, made by Marty Stouffer Productions, was provided by Animals, Animals.

Family Ties

Histocompatibility genes, which control the body's ability to distinguish self from nonself, determine whether a transplant recipient will reject a donor's tissue and are crucial to warding off infection. Recent evidence raises the possibility that histocompatibility genes may also play a role in kin recognition in a variety of species.

The latest findings come from Richard K. Grosberg and James F. Quinn of the University of California at Davis, who have shown that such genes probably control kin recognition in the tiny planktonic larvae of the sea squirt *Botryllus schlosseri*. They began by demonstrating that *B. schlosseri* larvae are in fact able to recognize their relatives. *B. schlosseri* is one of several marine invertebrates that are colonial. A colony is formed when a sexually produced larva leaves the pouchlike tunic enclosing its parental colony and swims to a rock or other firm material, where it attaches itself and metamorphoses into an essentially immobile individual called a zooid. This founder then reproduces by asexual budding, a process that results in a colony of genetically identical zooids sharing a common blood supply.

The investigators established a fertilized source colony whose offspring larvae would all carry a readily identifiable genetic marker. Then they placed the colony alone in the center of a circular asbestos-cement panel, put the panel in the sea, waited until larvae were likely to be born and mapped the locations of all newly founded *B. schlosseri* colonies on the panel. The maps revealed that larvae from the source colony, all of which were siblings, settled closer to one another than to nonsiblings that had migrated to the panel from the surrounding water. By means of mathematical modeling the workers also showed that the clustering was not solely the result of a general tendency to remain near the source colony; kin recognition exerted an independent influence.

Grosberg and Quinn suspected that histocompatibility genes mediated the kin recognition in part because they knew that recognition on the basis of histocompatibility genes would be an efficient way to identify a relative. In sea squirts, as in many species, the gene pool includes an extraordinary number of alleles, or variants, of histocompatibility genes; the likelihood is small that randomly associated nonrelatives will have identical alleles.

To study the influence of histocom-

patibility on kin recognition the workers put larvae whose histocompatibility alleles were known in petri dishes containing colonies whose histocompatibility alleles and relatedness to the larvae were also known. (In some cases the larvae were closely related to the colonies but were purposely bred to have different histocompatibility alleles.) Regardless of whether or not they were closely related to the petri-dish colonies, larvae that shared a histocompatibility allele with the resident colonies settled closer to the colonies than nonhistocompatible larvae did. In other words, relatedness per se had no effect on larval settlement patterns, indicating that histocompatibility is probably a crucial mediator of both aggregation and kin recognition in *B. schlosseri*.

The new results complement an earlier finding that mice are able to distinguish among individuals on the basis of histocompatibility: they preferentially mate with individuals whose histocompatibility alleles differ from their own, a practice that presumably helps them to avoid inbreeding (see "The Chemosensory Recognition of Genetic Individuality," by Gary K. Beauchamp, Kunio Yamazaki and Edward A. Boyse; SCIENTIFIC AMERICAN, July, 1985).

Grosberg and Quinn suggest that historecognition may also mediate kin recognition in many other organisms. It is conceivable, they say, albeit highly speculative, that historecognition even has a role in the kin recognition (and the consequent cooperation) often observed in mobile, social vertebrates, such as ground squirrels and prairie dogs.

PHYSICS

Of Atomic Bondage

The bonds that hold atoms together in a solid have been seen for the first time. The imaging feat has been accomplished by workers at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y., with the scanning tunneling microscope, an instrument that allows solid surfaces to be "seen" atom by atom.

Robert J. Hamers, Rudolf M. Tromp and Joseph E. Demuth describe their approach in *Physical Review Letters*. In operating the scanning tunneling microscope a voltage is applied between the sample being observed and the tip of a probe. When the tip is moved to within a few atomic diameters of the surface, electrons can "tunnel," or leak, through the intervening space. The tunneling phenomenon is a quantum effect; it is a direct consequence of the fact that particles behave like waves, so that their positions in space are smeared out.

The flow of electrons, which is registered as a current, depends sensitively on the distance between the surface and the tip. To image the bumpy terrain of an atomic surface, the tip is scanned back and forth over the sample. The tip's vertical height is continually adjusted to keep the current constant, and the changes in height are recorded to produce a contour map of the surface atoms.

To detect the bonds between the atoms, the IBM team periodically interrupted the scanning procedure, holding the probe tip still while they varied the applied voltage. From measurements of how the current changed they could construct a map showing where the electrons that bond surface atoms reside.

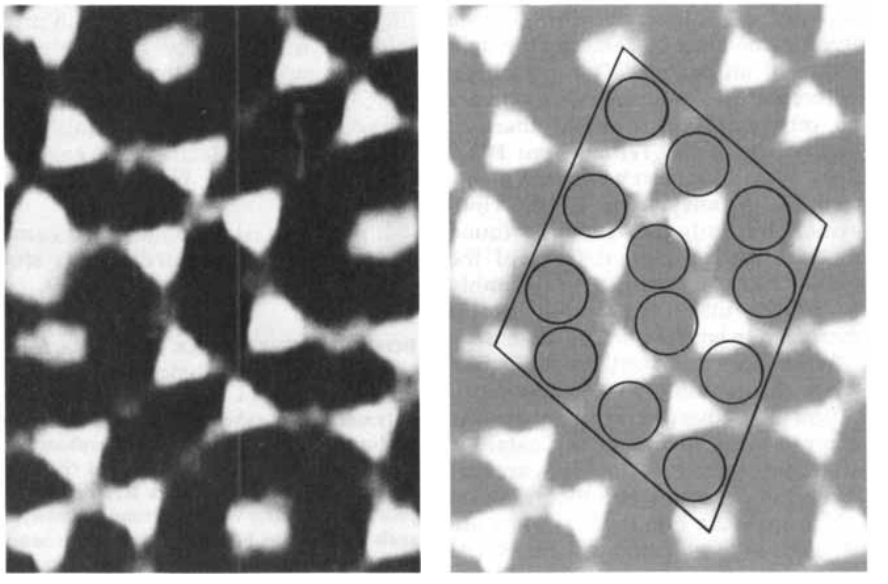
Quasar of a Different Color

Quasi-stellar objects, or quasars, are extremely bright celestial objects that lie billions of light-years from the earth. Even though their energy-producing core is only about the size of the solar system, quasars nonetheless outshine entire galaxies consisting of billions of stars. The 3,500 or so catalogued quasars emit most of their radiation in either the ultraviolet, the visible or the radio regions of the electromagnetic spectrum. Conspicuously absent had been a quasar that emits primarily infrared radiation.

An example of the "missing" quasar type has now been found. Charles A. Beichman, B. Thomas Soifer, George Helou, Thomas J. Chester and Gerry Neugebauer of the California Institute of Technology, Frederick C. Gillett of the National Optical Astronomical Observatory and Frank J. Low of the University of Arizona report the discovery of the infrared quasar, designated 13349 + 2438, in *Astrophysical Journal Letters*. The object was first detected as an infrared source in data returned by the *Infrared Astronomical Satellite* (*IRAS*) and was shown to be a quasar by subsequent ground-based observations with optical and radio telescopes.

An infrared quasar can be fitted neatly into a prevailing theory of quasar evolution, which holds that quasars arise from the collision of two galaxies. Infrared radiation such as that observed in the case of 13349 + 2438 could result from the heating of the thick interstellar dust clouds produced by the collision. Such dust clouds would obscure the radiation commonly associated with quasars until the dust dissipates, allowing the noninfrared radiation to escape and be detected on the earth.

As far as quasars go, 13349 + 2438 is relatively close to the earth: about a billion light-years away. This raises the possibility that other quasars even closer to the earth may now be in the process of forming. Many astronomers have thought quasar formation took place only during the early history of the universe, an idea consistent with the fact that all quasars seemed to be distant objects. If infrared quasars are found nearby, that hypothesis may have to be revised.



ATOMIC BONDS appear as bright spots between silicon atoms (*dark spots*). **The bonds reach up from the second layer of atoms in the sample. In the drawing (*right*) the diamond outlines a unit cell of silicon, which has 12 atoms and is repeated over the surface.**

RNA as an Enzyme

It was long thought that every cellular reaction is catalyzed by a protein enzyme. The discovery that RNA can cut, splice and assemble itself overturns the principle—and throws light on early evolution

by Thomas R. Cech

In a living cell the nucleic acids DNA and RNA contain the information needed for metabolism and reproduction. Proteins, on the other hand, are functional molecules: acting as enzymes they catalyze each of the thousands of chemical reactions on which cellular metabolism is based. Until recently it was generally accepted that the categories are exclusive. Indeed, the division of labor in the cell between informational and catalytic molecules was a deeply held principle of biochemistry. Within the past few years, however, that neat scheme has been overturned by the discovery that RNA can act as an enzyme.

The first example of RNA catalysis was discovered in 1981 and 1982 while my colleagues and I were studying an RNA from the protozoan *Tetrahymena thermophila*. Much to our surprise, we found that this RNA can catalyze the cutting and splicing that leads to the removal of part of its own length. If one could overlook the fact that it was not a protein, the *Tetrahymena* RNA came close to fulfilling the definition of an enzyme. One remaining difference was that enzymes operate on other molecules, rather than on themselves as the *Tetrahymena* RNA does. For this reason we coined a new term for the enzymelike RNA: "ribozyme." Recently, however, we found that a slightly different form of the same RNA can catalyze the assembly of RNA's other than itself, and is therefore an enzyme in the full sense.

What does the startling finding of RNA enzymes imply? The first implication is that one can no longer assume a protein lies behind every catalytic activity of the cell. It now appears that several of the operations that tailor an RNA molecule into its final form are at least in part catalyzed by RNA. Moreover, the ribosome (the organelle on which proteins are assembled) includes several molecules of RNA along with a variety of proteins. It may be that the RNA of the ribosome—rather than its protein—is the catalyst of protein synthesis, one of the most fundamental biological activities. RNA catalysis also has evolutionary implications. Since nucleic acids and proteins are interdependent, it has often been argued that they must have evolved together. The finding that RNA can be a catalyst as well as an informational molecule suggests that when life originated, RNA may have functioned without DNA or proteins.

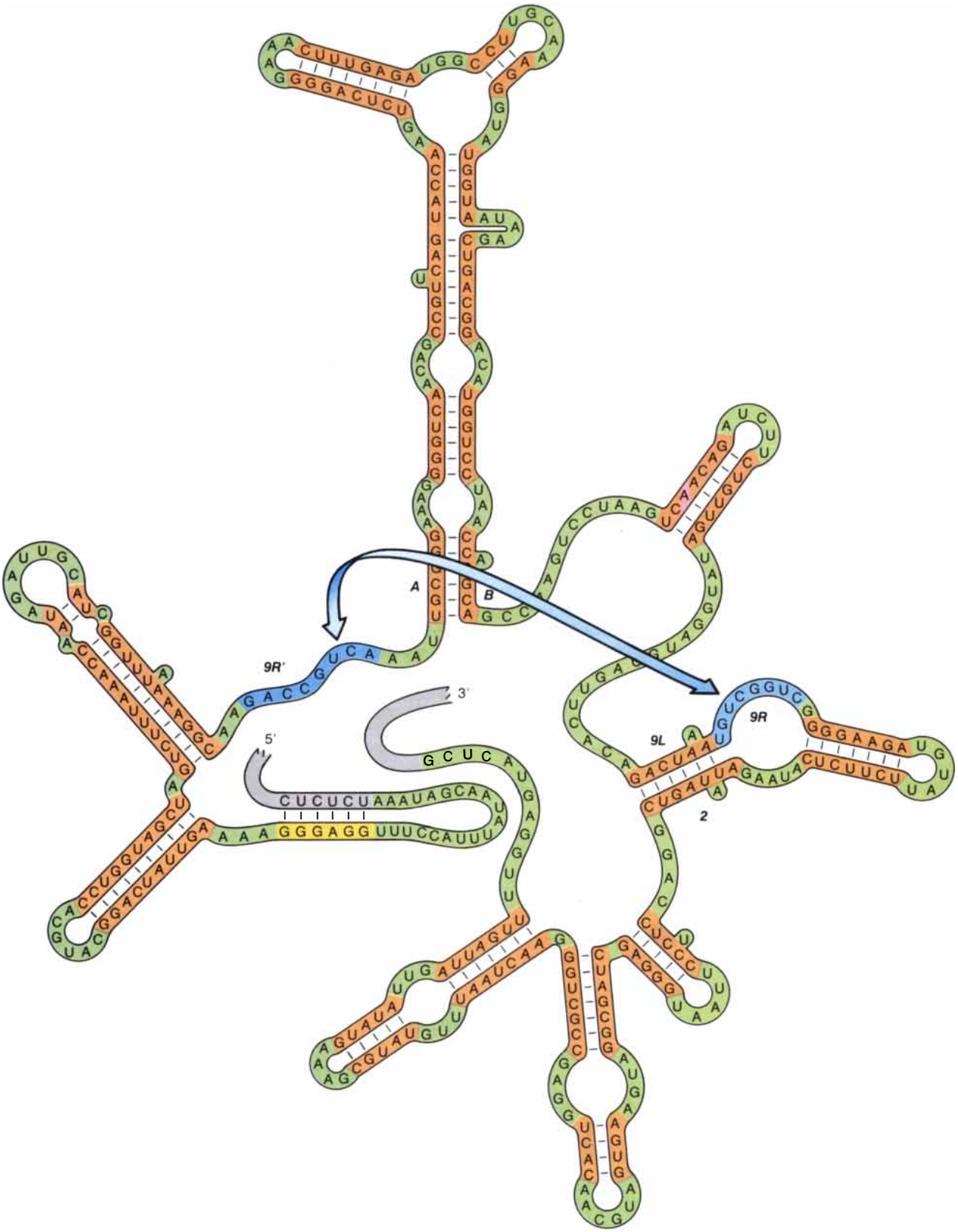
The Need for Catalysts

Why are catalysts so important in biological systems? The need for them stems largely from the nature of biological molecules themselves. The molecules that interact in cellular processes are generally quite stable; without outside intervention they are ponderously slow to react. Enzymes speed up these interactions and bring them into a time scale compatible with life. It is a job they do well. Each protein enzyme typically accelerates a biochemical reaction by a factor ranging from a million to a trillion. In doing so, the protein is a true catalyst: at the end of the reaction the enzyme emerges in the same form it had at the beginning.

It is easy to understand how it came to be generally accepted among students of the cell that every enzyme is a protein. In the decades following the purification of the first enzyme (urease, which helps to break down urea) by James B. Sumner of Cornell University in 1926, hundreds of enzymes were purified. In every case the enzyme was found to be a protein: a linked chain of amino acids folded into a specific three-dimensional shape. Along the way much was learned about how enzymes work. It was found that each enzyme catalyzes a single biochemical reaction or a group of closely related reactions. Such exquisite specificity is made possible by the folding of the amino acid chain, which endows the protein with a shape appropriate for participation in a given chemical interaction.

During the 1930's and 1940's, while the catalytic side of the cell was being deciphered, understanding of its informational aspect lagged. A major breakthrough came in 1944 when Oswald T. Avery and his colleagues at the Rockefeller Institute for Medical Research found that DNA is capable of conveying specific genetic information. Another breakthrough followed in 1953 when James D. Watson and F. H. C. Crick worked out the structure of DNA. These studies, along with many others, made clear what we now take for granted: that the genetic information resides in the sequence of

FOLDED STRUCTURE of an RNA from the protozoan *Tetrahymena* has a crucial role in the molecule's catalytic activity. That structure results in part from the capacity of a single-strand RNA to fold back on itself, forming short double-strand regions. RNA is a polymer whose subunits are nucleotides: adenosine (*A*), guanosine (*G*), cytidine (*C*) and uridine (*U*), each in its monophosphate form. *A* and *U* are complementary, or well suited to pairing, as are *C* and *G*. (*U* and *G* can also pair, although their interaction is weaker.) Hence if two stretches of an RNA molecule contain the appropriate nucleotides, they can join and establish double-strand segments (*orange*). Additional, long-range interactions—such as the one between the regions marked *9R* and *9R′* (*blue*)—lead to an even more complex three-dimensional shape. That shape helps the *Tetrahymena* RNA, which contains components of the organelle called the ribosome, to remove a portion of its own length called an intron. Most of the folded structure shown here is within the intron; only a small fraction of the remainder of the RNA is shown (*gray*). Models of the folding, all quite similar, have been worked out in several laboratories, including that of the author.



nucleotides in the DNA double helix. Each strand of a double-strand DNA molecule is a polymer whose subunits are the nucleotides adenosine (*A*), guanosine (*G*), thymidine (*T*) and cytidine (*C*), each in its monophosphate form. RNA is a single-strand polymer with a similar structure, although uridine (*U*) is substituted for thymidine.

Soon after the structure and significance of DNA began to be appreciated it was found that RNA has several key roles in the transfer of genetic information from DNA to protein. The first step in the expression of a gene is the transcription of the sequence of nucleotides in the DNA into the corresponding sequence of nucleotides in a messenger RNA (mRNA). The mRNA attaches itself to a ribosome, where it serves as a template for the assembly of a protein.

Carrying the genetic message is only one of the roles of RNA. The ribosome itself includes several ribosomal RNA (rRNA) molecules. Another small RNA called a transfer RNA (tRNA) helps to attach the amino acids to the growing protein chain in the correct order. By the late 1970's the roles of mRNA, rRNA and tRNA had long since been worked out, and it appeared that RNA was no longer a source of mystery. But the appearance was deceptive.

"Split" Genes

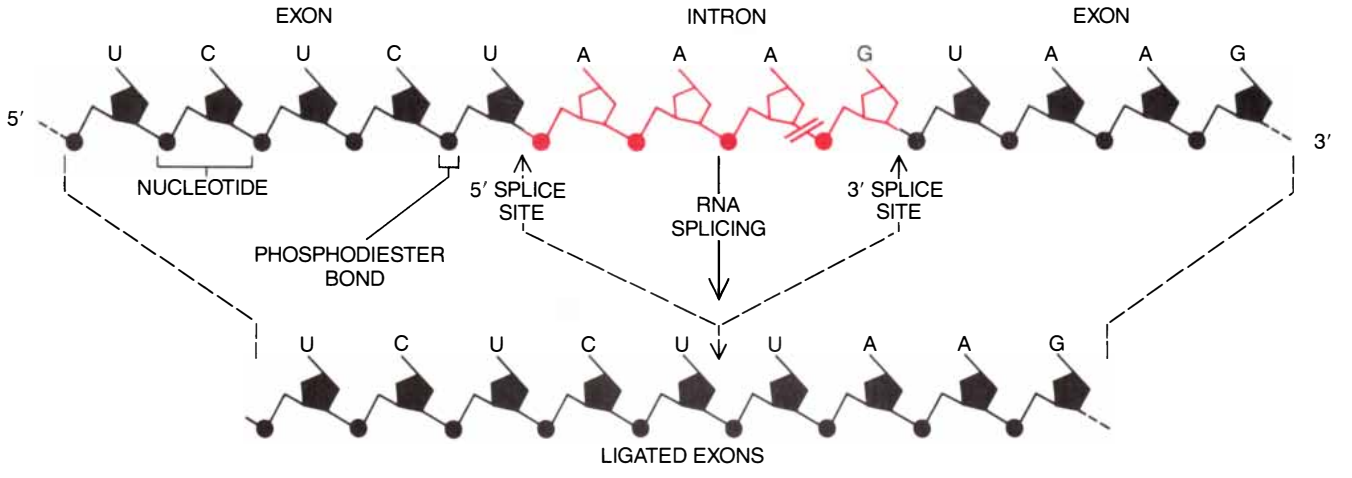
In 1977 two groups of research workers—Philip A. Sharp and his colleagues at the Massachusetts Institute of Technology and a group at the Cold Spring Harbor Laboratory—amazed themselves and the rest of the scientific world when they discovered "split" genes in higher organisms. The two groups found that the sequence of nucleotides in the DNA encoding a protein is not arranged continuously, as everyone had expected. Instead the coding sequences are interrupted by large stretches of noncoding DNA. The interrupting stretches were dubbed introns, or intervening sequences; the divided coding sequences were called exons. It did not take long to learn the fate of the introns: after the RNA is transcribed from the DNA, the introns are snipped out and the exons are spliced to form a continuous molecule.

The discovery of RNA splicing was tremendously exciting for several reasons. One reason was that splicing was found to take place in eukaryotes but not in prokaryotes, at least not in one well-studied prokaryote, the bacterium *Escherichia coli*. Eukaryotes are organisms—ranging from yeasts to human beings—that have nucleated cells. The prokaryotes—bacteria and certain algae—have cells that are not nucleated. It seemed plausible that the study of RNA splicing would illuminate the novel evolutionary potential of the eukaryotes, including their capacity for evolving specialized cells organized in multicellular organisms. It also seemed reasonable to suppose that RNA splicing would provide an important new form of regulation of gene expression.

In the late 1970's the problem of gene expression was being worked on in many laboratories, including my own. Specifically, we were studying the rRNA genes of the single-celled eukaryote *Tetrahymena thermophila*. In order to understand how those genes are expressed we needed to work out the series of cutting, splicing and other processing steps that yield the finished rRNA's. By 1980 we understood the outlines of the processing scheme. *Tetrahymena,* like all other eukaryotes, has four ribosomal RNA's, three of which are transcribed as a single unit of RNA. The resulting primary transcript is then cut, spliced and modified in other ways to produce the finished molecules.

One of the first processing steps, accomplished in the first few seconds after transcription, is the excision of an intron some 400 nucleotides long from the 6,400-nucleotide primary transcript; as the intron is excised the neighboring exons are ligated (spliced). In 1980 my colleague Arthur J. Zaug and I stumbled on evidence that the splicing reaction was taking place with high efficiency in test-tube systems containing mixtures of isolated *Tetrahymena* nuclei. By 1980 RNA splicing had been known for several years, but no mechanism for it had yet been defined, and so Zaug and I decided to devote part of our effort to finding out exactly how the *Tetrahymena* intron is excised.

The first step toward full understanding was to establish the minimum requirements for splicing in the test tube. Accordingly, unspliced pre-rRNA was mixed with extracts from *Tetrahymena* nuclei, which we assumed would serve as a source of the putative splicing enzymes. Along with these we added some small molecules such as salts and nucleotides. In their triphosphate form nucleotides often provide energy for cellular reactions,



INTRON IS REMOVED and the adjoining exons are spliced together to yield a functional RNA molecule. Each nucleotide in the RNA chain includes a five-carbon sugar molecule (shown as a pentagon), one of four bases (indicated by *A, G, C* and *U*) and a phosphate group (*shown as a circle*). A phosphodiester bond links adjacent nucleotides. Because the bond extends from the 5′ carbon on one sugar to the 3′ carbon on the next, the chain has a 5′ end and a 3′ end. When a molecule of RNA is processed, the phosphodiester bonds at both ends of the intron—called the 5′ splice site and the 3′ splice site—are broken. The splice sites are then joined.

and they were added in case external energy was needed to drive the splicing reaction. It turned out that some of the small molecules—notably magnesium ion and any of several forms of the nucleotide guanosine—were required for the reaction to proceed. To our great surprise, however, the nuclear extract containing the enzymes was not.

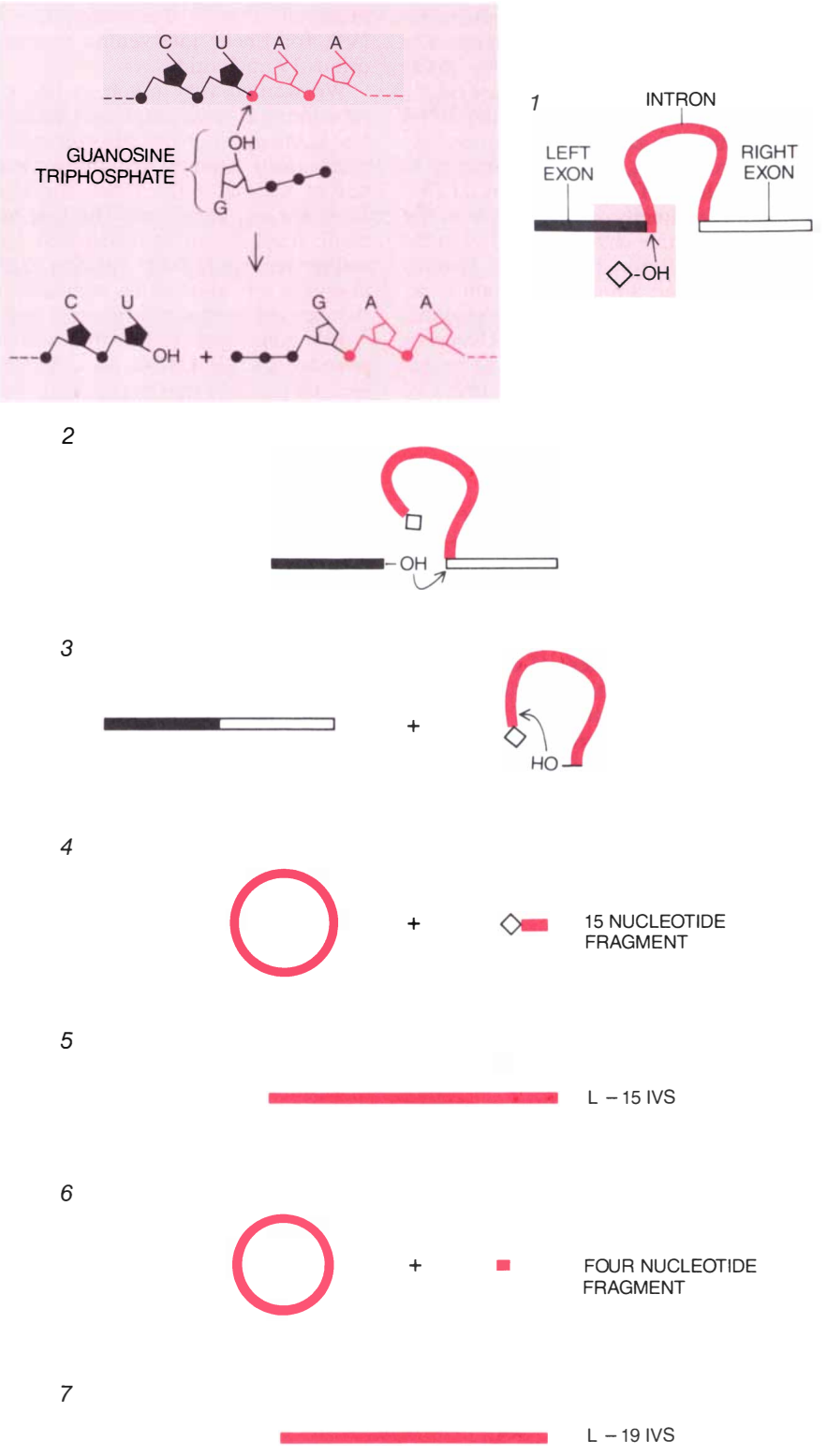
Is RNA an Enzyme?

We were forced to conclude either that the enzymatic activity came from a protein bound so tightly to the RNA that we were unable to strip it off or that the RNA was catalyzing its own splicing. Given the deeply rooted nature of the idea that all biological catalysts are proteins, the hypothesis of catalysis by RNA was not easy to accept. Yet soon we were confronted by another apparent example of RNA catalysis in experiments conducted by Paula J. Grabowski, who was then a graduate student in my laboratory. She found that the *Tetrahymena* intron also exists in a circular form, and that the linear molecule can be converted into the circular one by incubation in a solution containing magnesium ion—but no enzymes.

Once again the reaction appeared to be independent of all proteins, and hence of all enzymes, at least by the definition of an enzyme prevailing in 1981. Yet we still could not completely exclude the possibility that there was a protein enzyme hiding somewhere in the experimental system. The RNA used in our work was made in a fairly crude system consisting of isolated *Tetrahymena* nuclei. The possibility remained that an unusually hardy nuclear protein had become stuck to the RNA and was responsible for excising the intron, then converting it into a circle.

That possibility was eliminated by several experiments, the most telling of which entailed the synthesis of a truncated pre-rRNA transcript from a recombinant-DNA template. This system was entirely artificial: the DNA was grown in *E. coli,* then purified and transcribed into RNA by the addition of pure *E. coli* enzymes, which were subsequently removed from the solution. The resulting RNA had never been near a cell and therefore could not be contaminated by splicing enzymes. Nonetheless, it still spliced out its own intron. Not only that, the splicing was done at the same sites as in the intact cell—those sites having been defined by Joseph G. Gall of Yale University and his colleagues. There was no longer any doubt: the *Tetrahymena* intron is self-splicing.

It immediately struck us that in cer-



INTRON REMOVES ITSELF from the *Tetrahymena* ribosomal RNA (rRNA) precursor molecule with no assistance from protein enzymes. The cascade of reactions resulting in removal of the intron is unleashed by a free guanosine or guanosine triphosphate molecule (*shown as a diamond*). A hydroxyl group (OH) attached to the nucleotide "attacks" the phosphate at the 5′ end of the intron (*1*). The phosphodiester bond between the intron and the left exon is broken, and a new bond is formed between the guanosine and the intron (*inset*). This liberates a new hydroxyl group on the end of the left exon, which begins an attack at the 3′ end of the intron (*2*). The bond there is broken and the exons are ligated, or joined, freeing the intron (*3*). A similar reaction enables the intron to form a circle, snipping 15 nucleotides off its end in the process (*4*). The circle opens into a linear molecule (*5*) and then closes with the loss of four nucleotides (*6*). The final, reopened form is known as the L – 19 IVS for Linear Minus 19 Intervening Sequence (*7*).

tain fundamental ways self-splicing resembled the action of an enzyme. The reactions were accelerated by many orders of magnitude. They were highly specific. In addition, when the RNA was put in a solution that prevented it from folding, it lost its capacity to splice itself. This result showed that the three-dimensional structure of the polynucleotide chain is essential to the splicing activity, just as the specific folding of an amino acid chain is required for the activity of an enzyme. Yet unlike an enzyme, the RNA was operating on itself, and it was to acknowledge this difference that we coined the term ribozyme.

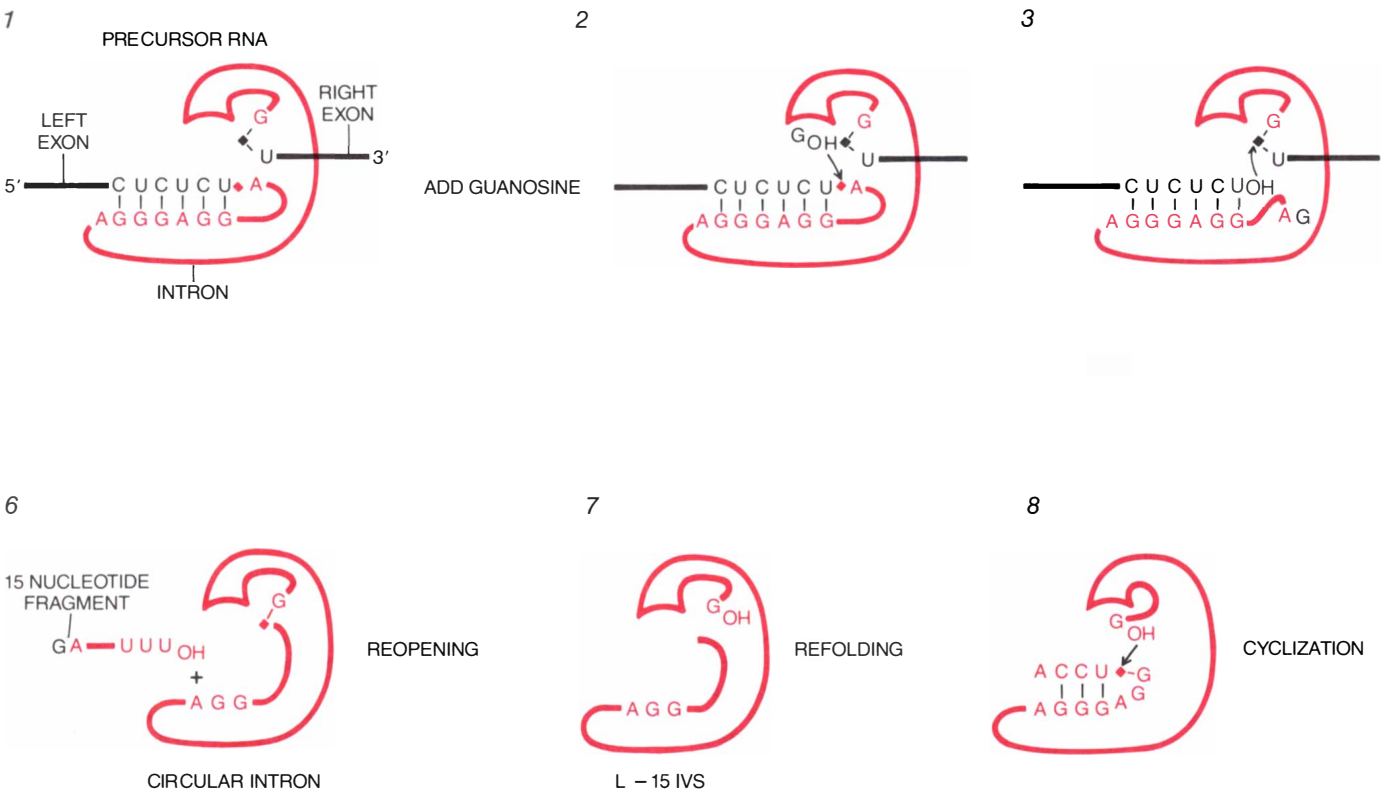
A Catalytic Cascade

Self-splicing turned out to be only one step in a cascade of reactions that the *Tetrahymena* RNA carries out on its own. After the intron is removed it immediately converts itself into the circular form discovered by Grabowski. In the process of cyclizing itself, the intron, initially 414 nucleotides long, loses a stretch of 15 nucleotides. After a brief interval it opens up again into a linear form. The linear molecule then forms a second circle, this time losing four nucleotides. The circle again reopens, yielding a linear molecule that is designated L – 19 IVS, for linear intervening sequence minus 19 nucleotides.

We wanted very much to find out how the intron accomplishes these remarkable gyrations. Some clues to the mechanism began to accumulate even before we had proved that the reactions are self-catalyzed. The first hint came from the observation that guanosine is required for splicing. Zaug analyzed the nucleotide sequence of the excised intron and found it began with a guanosine residue that was not encoded by the DNA. As a result it seemed possible that the splicing reaction required guanosine not as an energy source but as a monomer to be added to the intron. To test that possibility I mixed some radioactively labeled GTP (guanosine triphosphate, the form employed as an energy source in many cellular reactions) with unlabeled pre-rRNA. The GTP was not utilized for its energy. Instead, when the splicing was finished, the radioactive GTP appeared on one end of the intron.

What part, then, was guanosine playing in the splicing reaction? We postulated that GTP (or any of several other forms of guanosine, which worked equally well) was acting as an "attacking group" to help break the RNA chain at the correct site. Cleaving a strand of RNA requires the breaking of its backbone, which consists of two kinds of molecules in an alternating pattern: phosphate groups and five-carbon sugars. (The nitrogenous bases that vary from nucleotide to nucleotide, thus encoding the genetic information, are attached to the backbone, one per sugar.) The linkage between the phosphate group and the sugar—known as a phosphodiester bond—must be sundered for the RNA chain to part. By attacking the phosphodiester bond at one end of the intron, guanosine initiates the self-splicing reaction.

Starting with what we had found out about the guanosine attack, we went on to work out the overall scheme of the splicing reactions. The process begins when a free guanosine (or GTP) molecule encounters the intron RNA and is temporarily held in a molecular cradle called a binding site formed by a portion of the intron. The binding site positions a hydroxyl group (OH) on the guanosine in a favorable position for attacking the phosphate group at the junction between the exon and the 5′ end of the intron. (Each phosphodiester bond extends from the 5′ carbon of one sugar to the 3′ carbon of the next, and as a result the RNA



SEVERAL CATALYTIC STRATEGIES aid the *Tetrahymena* intron in its repeated phosphodiester exchanges. The nucleotide sequence *GGAGGG* (such sequences are conventionally given in the 5′-3′ direction) near the end of the intron binds the sequence *CUCUCU* at the end of the left exon (*1*). At the same time the phosphate group at the 5′ end of the intron is activated (*colored diamond*). When guanosine is added, the intron binds it in a position favoring its attack on the exposed, activated phosphodiester bond (*2*). Once the bond has been severed, the phosphate at the other end of the intron, which has also been activated (*black diamond*),

strand is directional: one end is 5′ and the other is 3′.)

At the same time a second binding site holds the end of the exon, exposing the phosphate linkage to the attack of the guanosine. Although the precise structure of the binding site for the guanosine is not as yet known, the binding site for the end of the exon has been identified. It turns out to be a string of nucleotides in the intron with the sequence *GGAGGG*. Those six nucleotides pair with a string of six pyrimidines (*CUCUCU*) at the 5′ end of the exon. (The nitrogenous bases of the RNA come in two forms: *A* and *G* are purines, *C* and *U* are pyrimidines.) By simultaneously grasping the six pyrimidines and the guanosine, the intron brings the reactants into position for guanosine to begin its attack.

The object of guanosine's attack is the bond between the final *U* in the string of pyrimidines and the adjacent nucleotide, which is an *A*. The guanosine breaks the *U-A* link and inserts itself between them, forming a new bond between itself and the *A*. Such a swapping of phosphodiester bonds is called a transesterification reaction. As the guanosine is added to the end of the intron, the 3′ end of the exon with its terminal *U* dangles free. The *U* dangles only for a moment. Quickly a second transesterification joins the exons, excising the intron. A third such reaction converts the intron into a circle and, after the circle has been broken, a fourth restores it to a slightly smaller circular form.

Strategies of the Ribozyme

Working out the mechanisms of the self-splicing (in collaboration with my colleagues, including Brenda L. Bass, Francis X. Sullivan, Tan Inoue and Michael D. Been) further emphasized the similarity between the ribozyme and an enzyme. A common catalytic strategy of enzymes is to bind two substrates (reactants) in close proximity and at a special orientation to each other, thereby facilitating their reaction. The RNA intron employs the same strategy when it binds guanosine and the six pyrimidines.

The resemblance did not end there. We already knew that the folded structure of the intron is essential to its activity, just as the three-dimensional form of a protein is. One reason the folding is essential is that it creates the sites for the binding of the guanosine and the stretch of pyrimidines. As the detailed mechanism of the interaction was revealed, we found that the folded structure of the intron also activates the phosphate group at each reaction site, bringing it into a state favorable for cleavage. Evidence for this conclusion came from the observation that even in the absence of guanosine the intron undergoes a slow cleavage reaction at the 3′ splice site and—to a lesser extent—at the 5′ splice site as well. In the absence of guanosine the hydroxyl ions in the solution act as the attacking group.

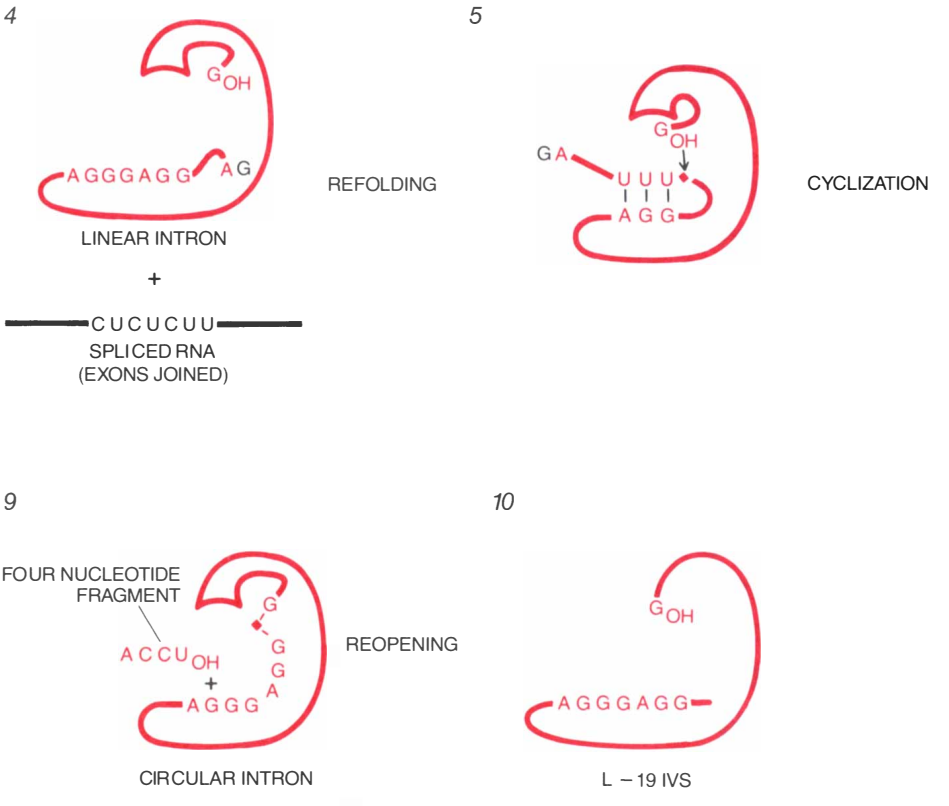
Cleavage without guanosine is slower than cleavage with guanosine. That is not surprising, since hydroxyl ion cannot manage the binding interactions that position guanosine for optimum attack. What is significant, however, is that even without guanosine the cleavage takes place at the same sites as those in the full splicing reaction. Somehow the folded structure of the intron favors the cutting of the RNA at just the sites that form the ends of the excised intron. Even without the attacker that serves in the biological reaction (guanosine) it is at those sites that the cuts are made.

Precisely how the folding activates the phosphates at the splice sites is not known, but that should not keep us from appreciating how effective a strategy the activation is. RNA is ordinarily quite stable; its phosphodiester bonds break very slowly under uncatalyzed conditions. At the splice sites the folded structure of the intron accelerates the rate of cleavage by a factor of 10 billion. To put that ratio in a more accessible form: the folded structure reduces the time frame of the reaction from 19,000 years to one minute.

What Is the Core Structure?

What is the folded shape that is capable of having such a potent effect? The answer draws on findings that range far afield into the relations between the *Tetrahymena* RNA and similar molecules in many other species. The other species entered the picture because by 1982 an overall classification scheme for RNA-splicing reactions was beginning to emerge. Initially the scheme was merely a categorization of various RNA's according to similarities in the nucleotide sequences of their introns. The classification included four groups. One group was made up of the transcripts processed in the nucleus to yield tRNA's. A second was made up of the molecules, also processed in the nucleus, that yield mRNA's.

The other categories—designated Group I and Group II—include RNA's from more diverse sources. Group I includes an assortment of rRNA, mRNA and tRNA introns, many of them from mitochondria and chloro-



is attacked by a hydroxyl group attached to the left exon (*3*). The exons are thereby joined and the intron is released (*4*). The intron then refolds, and the sequence *UUU* near its own 5′ end is positioned for the attack of the guanosine at its 3′ end (*5*). The 15 nucleotides lost as the loop closes are those preceding the activated phosphate at the point of attack (*6*). After the circle opens (*7*) it recyclizes (*8, 9*) before opening again (*10*).

Once again, we've been placed on a pedestal.

"Making it to the top is generally much easier than staying at the top once you have made it. Unless you are Honda . . ." *Motor Trend.*

"Hope doesn't burn brighter than it does for the unveiling of an all-new Accord. Honda Accords are great automobiles, and everybody knows it." *Car and Driver.*

"It's improved in just about every way, objectively as well as subjectively." *Road & Track.*

"With the new Accord, and the LXi in particular, Honda has solidified its position among the world's truly fine sports sedans." *Motor Trend.*

"There is a solid, jiggle-free feeling from the new suspension. It makes the Honda Accords

feel almost exactly like Audi's 18-inch longer top line 5000. The ride approaches that of Mercedes 190, which features a much more complex five-link rear suspension." *Autoweek.*

"The engine seems to hum even more sweetly than before. Cruising is hushed. The steering is deliciously accurate. And the shifter carves to perfection." *Car and Driver.*

"Cruising the high-banked oval at the Accord's top speed, passengers in the front and rear carried on hushed conversation in a dead-flat acoustic environment. Wind noise was minimal and road noise was dampened effectively." *Motor Trend.*

"Que será, será." *Anonymous Accord LXi owner.* HONDA

plasts (cytoplasmic organelles containing their own DNA, RNA and ribosomes). Some 30 Group I introns have been identified so far (there may be many more), and the species in which they have been found range from protists to fungi to higher plants such as corn and beans but do not include animals. Group II RNA's, which come mainly from fungal mitochondria and plant chloroplasts, appear to be less common.

Group I introns were originally defined as fungal mitochondrial introns that have four nucleotide sequences in common, each sequence about 10 nucleotides long. It soon became clear that those sequences have a strong bearing on the mode of action of the Group I introns, and in particular on the three-dimensional form that underlies splicing. The first hint of the relation between sequence and function came in 1982, when several research groups found that the *Tetrahymena* intron, which is a nuclear molecule, shares the same conserved sequence elements.

The finding of common sequences was surprising, because the RNA's in question come from quite disparate sources. (Fungi and protozoans are not thought to be closely related, and mitochondrial nucleic acids usually differ considerably from their nuclear counterparts.) The preservation of a nucleotide sequence across a broad range of species often implies that the sequence has a significant functional role, which must be maintained in the face of evolutionary divergence. Taking into account what was known about the *Tetrahymena* intron, it seemed reasonable to think the role of the conserved sequences in the Group I introns lay in forming the three-dimensional structure that directs self-splicing.

Models of the Core

Ideas about what the core structure might be came quickly from several sources, and the agreement among them imparted a certain confidence in their validity. François Michel and Bernard Dujon of the Center for Molecular Genetics at Gif-sur-Yvette in France and R. Wayne Davies of the Victoria University of Manchester and his colleagues proposed independently that the conserved sequence elements help to fold the Group I introns into three-dimensional structures that are similar, at least in a core region.

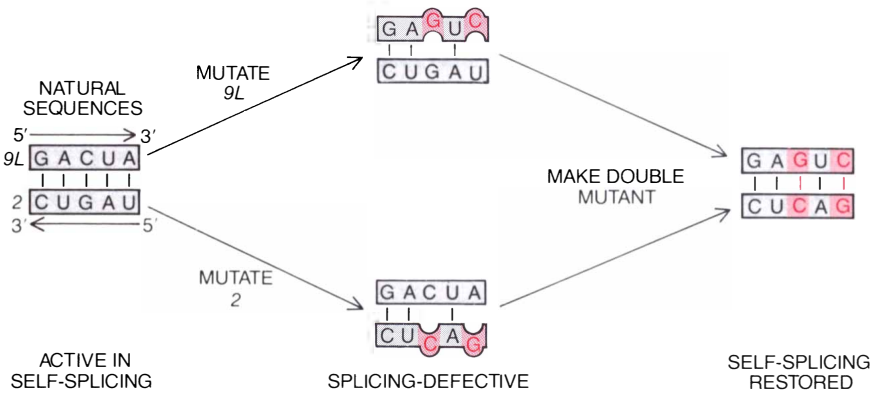
One way the shared sequences determine the core structure is by binding to each other. Among the four nucleotides of RNA, two pairs (*A* and *U; C* and *G*) are complementary, or particularly adept at forming the fairly weak type of chemical bond known as a hydrogen bond. Hydrogen bonds between complementary nucleotides provide the basis for the pairing of the two strands of the DNA double helix, and, in a similar fashion, two stretches of RNA can form a helical region if they contain complementary nucleotides. By pairing with one another, the conserved sequences might determine that the intron folds up into a particular shape, and that is what Michel and Davies proposed after analyzing the nucleotide sequences.

At the same time our group was taking a somewhat different, biochemical approach to the problem. Combining structural data obtained by N. Kyle Tanner, a graduate student in my laboratory, with a computer program developed by Michael Zuker of the National Research Council of Canada, we devised a model for the structure of the intron. Our model agreed almost perfectly with the ones Michel and Davies had worked out. The correspondence reassured us that the models are fundamentally correct, and presumably the shape they define has a direct role in self-splicing.

Other evidence suggests the hypothesis is correct. Geneticists working on mitochondrial gene expression in yeast have found a class of mutations (changes in the nucleotide sequence of a gene) that alter the short conserved sequences in Group I introns and prevent splicing in the living cell. More recently we and the Davies group have disrupted the conserved sequences in the *Tetrahymena* intron and found that self-splicing was thereby inhibited. Such studies clearly imply that the shared sequences contribute to a shape favoring the intron's catalytic maneuvers. Yet significant questions remain, because the details of the full three-dimensional structure are still being worked out. It may soon be clearer how the folded shape of the core enables the intron to bring the correct substrates together and encourages them to react.

As the relation between the core structure and self-splicing was coming into view, several discoveries combined to show that the *Tetrahymena* intron is far from unique. This was reassuring, because it helped to persuade us that our findings were correct, but in one sense it was not really surprising. After all, if the *Tetrahymena* intron and the Group I mitochondrial introns have a similar structure, and that structure is required for self-splicing, one might expect that the Group I introns also possess the capacity for self-splicing.

Recently several groups have confirmed that expectation. Gian Garriga and Alan Lambowitz of St. Louis University demonstrated that a mitochondrial mRNA intron from the fungus *Neurospora crassa* is self-splicing. Henk Tabak and his colleagues at the University of Amsterdam found several self-splicing mRNA and rRNA introns from yeast mitochondria. In all these instances guanosine is required and the RNA is spliced by the same



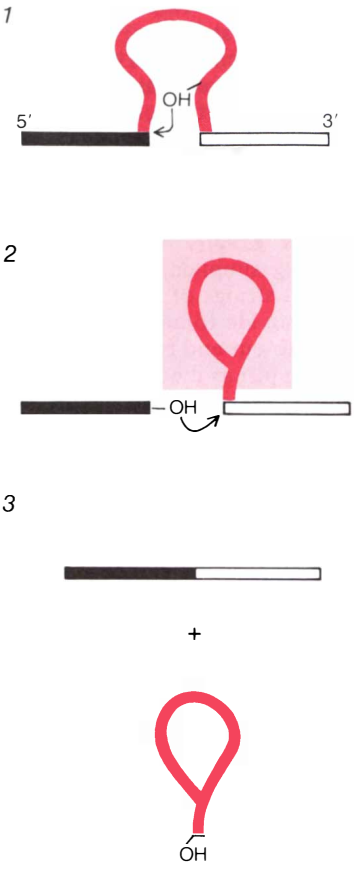
SIGNIFICANCE OF BASE PAIRING for the catalytic activity of the *Tetrahymena* intron was demonstrated in experiments that destroyed the intron's capacity to form double-strand regions. Among the sequence elements that have a particularly significant role in determining the folded structure of the intron's core region (*shown in the illustration on page 65*) are the ones designated *9R′*, *A*, *B*, *9L*, *9R* and *2*. In the fully functional intron, *9L* pairs with *2* to form a double-strand region. When mutations are introduced into *9L* or *2* that change the nucleotide sequences and prevent normal pairing (as was recently done by the author in collaboration with John M. Burke of Williams College and Burke's colleagues), the intron is no longer capable of splicing itself out of the rRNA precursor. Combining the two sets of mutations reestablishes pairing and restores the catalytic activity of the intron. The functional significance of sequences *9R′*, *A*, *B*, *9L* and *2* is further emphasized by the fact that they have been conserved through evolution in all introns related to that of the *Tetrahymena* pre-rRNA, which make up a category called Group I.

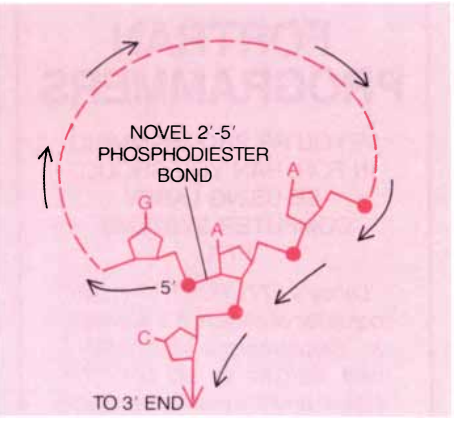
mechanism as in *Tetrahymena*. Remarkably, it has since been shown that an RNA from a bacterial virus—a very different category of biologic object—can manage the same feat. Marlene Belfort, Frank Maley and their colleagues at the New York State Department of Health in Albany found that an RNA of the bacteriophage *T4* splices itself in *E. coli*.

The discovery by Belfort and Maley extended the range of the Group I mechanism quite far—from bacterial viruses through higher organisms. Newer findings have extended self-splicing into a new category altogether: the Group II introns. Craig L. Peebles of the University of Pittsburgh, working with Philip S. Perlman of Ohio State University, and Leslie Grivell and his colleagues at the University of Amsterdam this year reported finding a Group II mitochondrial mRNA precursor that can remove its own intron. The Group II self-splicing mechanism is not the same as that of the Group I molecules. For one thing, it does not require guanosine, or indeed any free nucleotide. In addition, the process entails the formation of an intermediate structure not seen in the *Tetrahymena* intron or its Group I cousins: the lariat.

The lariat is a loop of RNA formed by the intron as it excises itself from a precursor RNA transcript. The loop is not formed by the joining of the two ends of the intron. Instead one end forms a bond with a nucleotide a short distance from the other end of the intron, leaving a stub of RNA extending beyond the loop. The structure formed in this way bears a passing resemblance to the rope twirled by cowboys, hence the name lariat.

Intriguingly, the lariat found in the Group II splicing process had already been found as a product of the splicing of nuclear pre-mRNA. Such splicing requires proteins, and it had always been assumed that nuclear pre-mRNA splicing is carried out by protein enzymes rather than by a self-splicing mechanism. The finding of the lariat in self-splicing Group II introns, however, suggests ribozymes may also have a role in the splicing of the nuclear messengers. Moreover, Group II introns may represent an evolutionary link between the nuclear pre-mRNA's and the self-splicing RNA's of Group I.





LARIAT is a loop formed by a category of introns—designated Group II—as they are removed from their RNA molecules. Some Group II introns are self-splicing, but their splicing does not require guanosine. Instead one of the many 2′ hydroxyl groups on the intron attacks the 5′ splice site (*1*). The subsequent reaction joins the 5′ end of the intron not to the 3′ end but to a point a short distance away, yielding a branched structure with a loop: the lariat (*2*). The branching is accomplished by the formation of a novel 2′-5′ phosphodiester bond that enables one adenosine nucleotide to form phosphodiester links with three other nucleotides rather than with the usual two (*inset*). Ligation of the exons frees the lariat from the remainder of the RNA (*3*).

True Catalysis

Although it was by now established that self-splicing RNA had many enzymelike properties, one difference persisted: the fact that the ribozyme was operating on itself rather than on another molecule. Since it altered itself in the course of the self-splicing reaction, it could not be considered a catalyst in the strict sense. For some investigators that lingering difference between ribozymes and enzymes put them in distinctly different categories. Only two years after the discovery of RNA self-splicing even that distinction disappeared.

In 1983 Sidney Altman and his colleagues at Yale University were collaborating with a group led by Norman R. Pace at the National Jewish Hospital in Denver to examine the action of ribonuclease *P*. Ribonuclease *P* is a tRNA-processing enzyme found both in bacteria and in higher cells; it is an unusual enzyme in that it contains RNA and protein in a single package. Altman and others had previously shown that under cellular conditions the RNA and the protein are both needed for the nuclease to carry out its work of cutting the tRNA precursor at a specific point. Altman and Pace found, however, that in the test tube the RNA subunit alone can cut the pre-tRNA molecule at the correct spot, whereas the protein has no such capacity.

Early the following year Altman reported the final proof that the RNA component of ribonuclease *P* was the catalytic subunit. A recombinant-DNA template was exploited to transcribe the RNA subunit of the ribonuclease, and it was shown that the RNA made from the template could catalyze the accurate maturation of tRNA precursors. By eliminating the possibility that catalysis could be due to a contaminating protein, this experiment proved that—in the test tube at least—the RNA subunit acts as an enzyme in the complete sense. Presumably the protein, which by itself lacks catalytic activity, helps the RNA to function under the conditions prevailing in the living cell.

Recently Zaug and I found that a shortened form of the *Tetrahymena* intron also has the capacity to act as a true enzyme. The experiment leading to that conclusion was stimulated by thinking about how the intron converts itself into a circle. Both of the intron cyclization reactions have the same general form: the 5′ end of the intron is bound and clipped off while a transesterification reaction joins the remainder of the intron into a loop. This reaction is in essence identical with the one leading to the excision of the intron, and as in the excision the RNA being clipped off must contain a stretch of pyrimidines that can attach itself to the binding site.

After the intron has twice been cyclized it can no longer transform itself into a circle, chiefly because the 19 nu-

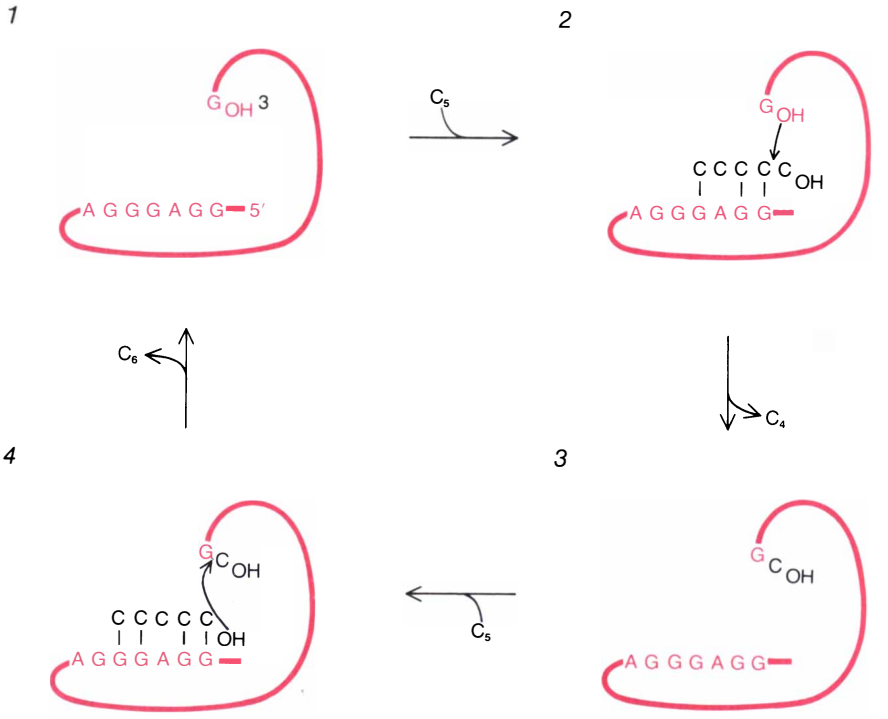
cleotides cleaved from its 5′ end contain the only strings of pyrimidines available for binding. The shortened molecule, however, has not necessarily lost its catalytic activity. We reasoned that if the truncated intron—the L − 19 IVS, in fact—were supplied with a new substrate, it might then act as a true enzyme.

That is exactly what transpired when Zaug supplied the L − 19 IVS with short strings of pyrimidines—*C*'s in this case. He found that the intron was able to cut and rejoin the pyrimidine chain just as it cuts and rejoins itself when the right sequences are present at its 5′ end. The L − 19 IVS emerges unscathed from each round of reactions, ready to cut and rejoin another string of pyrimidines. Therefore it satisfies the definition of a true catalyst.

Although we had predicted the enzymatic activity of the shortened intron, we were still startled by how good an enzyme it is. The L − 19 IVS can cut and splice the chain of *C*'s about 100 times an hour. Even though the truly enzymatic reaction requires two separate molecules to find each other before they can react, each reaction occurs just as fast as the self-splicing reaction.

RNA Can Assemble RNA!

There was another surprise. The repeated cutting and splicing lengthened the string of pyrimidines: the L − 19 IVS was synthesizing a polymer made of *C*'s. That was a finding with profound implications, because RNA polymerization is a prerequisite for the duplication of RNA genes, which has long been considered fundamental to the origin of life. The discoveries of self-splicing and the enzymatic activity of ribonuclease *P* had already been thought by some to resolve the chicken-and-egg problem of early evolution (concerning whether protein or nucleic acids came first) in favor of RNA. The discovery that the *Tetrahymena* intron has a rudimentary capacity to synthesize RNA puts that conclusion on firmer ground. It offers evidence that RNA can, in principle at least, catalyze its own replication. Now it seems reasonable to imagine that the first step toward life was the replication of RNA without the assistance of any functional proteins.



TRUE ENZYMATIC ACTIVITY of an intron—catalyzing an operation on another molecule rather than on itself—was first demonstrated with the *Tetrahymena* L − 19 IVS RNA. The intron-enzyme lacks the sequence near the 5′ end that is ordinarily bound to the binding site (*GGAGGG*) and thereby exposed to attack by the guanosine at the 3′ end of the intron. The binding site, however, is still present (*1*), and it can bind other molecules with complementary sequences, such as a chain of five cytosines (*2*). The ensuing reaction links the final cytosine to the guanosine at the 3′ end of the intron (*3*). Other chains of cytosine are present, one of which binds to the intron and begins an attack on the bond between the cytosine and the guanosine (*4*). The result is a chain of six cytosines, which can serve as the substrate for adding another cytosine (not shown). Chains up to 30 nucleotides long have been assembled by this method in the author's laboratory.

Having wandered back into the prebiotic past, it is fun to peer into the future a bit and speculate about where the next examples of RNA catalysis might be found. In all known examples the substrate for the RNA enzyme has been RNA: another part of the same molecule, a different RNA polymer or a single nucleotide. This is probably not accidental. RNA is well suited to interacting with other RNA's, but it is more difficult to envision RNA forming a good active site with other biologically significant molecules such as amino acids or fatty acids. Hence I expect that future examples of RNA catalysis will also entail RNA as the substrate.

Two possibilities come to mind. One involves the small nuclear ribonucleoprotein particles (snRNP's, or, as they are often called, "snurps") that are required for many RNA-processing operations in the nucleus, including splicing pre-mRNA. Each snurp consists of an RNA and several proteins. It is possible that the snurps simply mark the sites on mRNA precursor molecules where traditional protein enzymes cut and splice, but I suspect they have a more direct role in catalysis.

The other possibility is the ribosome, which includes 55 or more proteins in addition to the rRNA's. Some of the substrates with which the ribosome interacts are also RNA's, specifically mRNA and tRNA. Research in many laboratories—most notably those of Carl R. Woese of the University of Illinois at Urbana-Champaign and Harry F. Noller, Jr., of the University of California at Santa Cruz—has shown that in the course of evolution the structure of the rRNA's has been stringently conserved. Their work, combined with the discovery of mutations in the rRNA that affect protein synthesis, may indicate that the ribosomal RNA has a direct catalytic role in protein synthesis.

The conclusion that protein synthesis, a fundamental biosynthetic activity, is catalyzed by RNA would be a final blow to the idea that all cellular function resides in proteins. Of course, it may not be so; the ribosome may be such an intimate aggregation of protein and nucleic acid that its catalytic activity cannot be assigned exclusively to either component. Yet whether or not the synthetic activity of the ribosome can be attributed to the rRNA, a fundamental change has taken place in biochemistry in the past five years. It has become evident that, in some instances at least, information-carrying capacity and catalytic activity inhere in the same molecule: RNA. The implications of this dual capacity are only beginning to be understood.

The Higgs Boson

It could give mathematical consistency to the standard model—the theory that describes the interactions of fundamental particles. The search for the elusive particle will require new accelerators

by Martinus J. G. Veltman

The truly fundamental problems of physics can always be explained in simple terms without the help of complicated equations or mathematical arguments. At least this was once told to me by Victor F. Weisskopf, an eminent physicist who often engages in such explanations, and he may very well be right. It certainly holds for a proposed but undiscovered particle called the Higgs boson and the so-called Higgs field associated with it.

The Higgs boson, which is named after Peter W. Higgs of the University of Edinburgh, is the chief missing ingredient in what is now called the standard model of elementary processes: the prevailing theory that describes the basic constituents of matter and the fundamental forces by which they interact. According to the standard model, all matter is made up of quarks and leptons, which interact with one another through four forces: gravity, electromagnetism, the weak force and the strong force. The strong force, for instance, binds quarks together to make protons and neutrons, and the residual strong force binds protons and neutrons together into nuclei. The electromagnetic force binds nuclei and electrons, which are one kind of lepton, into atoms, and the residual electromagnetic force binds atoms into molecules. The weak force is responsible for certain kinds of nuclear decay. The influence of the weak force and the strong force extends only over a short range, no larger than the radius of an atomic nucleus; gravity and electromagnetism have an unlimited range and are therefore the most familiar of the forces.

In spite of all that is known about the standard model, there are reasons to think it is incomplete. That is where the Higgs boson comes in. Specifically, it is held that the Higgs boson gives mathematical consistency to the standard model, making it applicable to energy ranges beyond the capabilities of the current generation of particle accelerators but that may soon be reached by future accelerators. Moreover, the Higgs boson is thought to generate the masses of all the fundamental particles; in a manner of speaking, particles "eat" the Higgs boson to gain weight.

The biggest drawback of the Higgs boson is that so far no evidence of its existence has been found. Instead a fair amount of indirect evidence already suggests that the elusive particle does not exist. Indeed, modern theoretical physics is constantly filling the vacuum with so many contraptions such as the Higgs boson that it is amazing a person can even see the stars on a clear night! Although future accelerators may well find direct evidence of the Higgs boson and show that the motivations for postulating its existence are correct, I believe things will not be so simple. I must point out that this does not mean the entire standard model is wrong. Rather, the standard model is probably only an approximation—albeit a good one—of reality.

Even though the only legitimate reason for introducing the Higgs boson is to make the standard model mathematically consistent, much attention has been given to the conceptually easier proposal that the particle generates the masses of all the fundamental particles. I shall therefore begin with that topic.

Central to an understanding of how the Higgs boson would generate mass is the concept of a field. A field is simply a quantity, such as temperature, defined at every point throughout some region of space and time, such as the surface of a frying pan. In physics the term "field" is usually reserved for such entities as the gravitational field and the electromagnetic field. Fields generally make themselves felt by means of the exchange of a mediating particle; the particle that mediates the electromagnetic field, for example, is the photon, or quantum of light. The mediating particles of the gravitational field, the weak field and the strong field are respectively the graviton (which has not yet been detected), three weak vector bosons, called the W^+, W^- and Z^0 particles, and eight gluons. In a somewhat analogous way the Higgs boson is the mediating particle of the proposed Higgs field.

It is now assumed that there is a constant Higgs field throughout all space, that is, the vacuum of outer space is not empty but contains this constant field. The Higgs field is thought to generate mass by coupling to particles. Depending on the coupling strength, a particle in space has a certain potential energy. By Einstein's famous equation, $E = mc^2$ (energy equals mass multiplied by the square of the speed of light), the coupling energy is equivalent to a mass. The stronger the coupling, the greater the mass.

The way particles are thought to acquire mass in their interactions with the Higgs field is somewhat analogous to the way pieces of blotting paper absorb ink. In such an analogy the pieces of paper represent individual particles and the ink represents energy, or mass. Just as pieces of paper of differing size and thickness soak up varying amounts of ink, different particles "soak up" varying amounts of energy, or mass. The observed mass of a particle depends on the particle's "energy-absorbing" ability and on the strength of the Higgs field in space.

What are the characteristics of the proposed Higgs field? In order to endow particles with mass, the Higgs field, if it exists, would have to assume a uniform, nonzero value even in the vacuum. Moreover, the Higgs

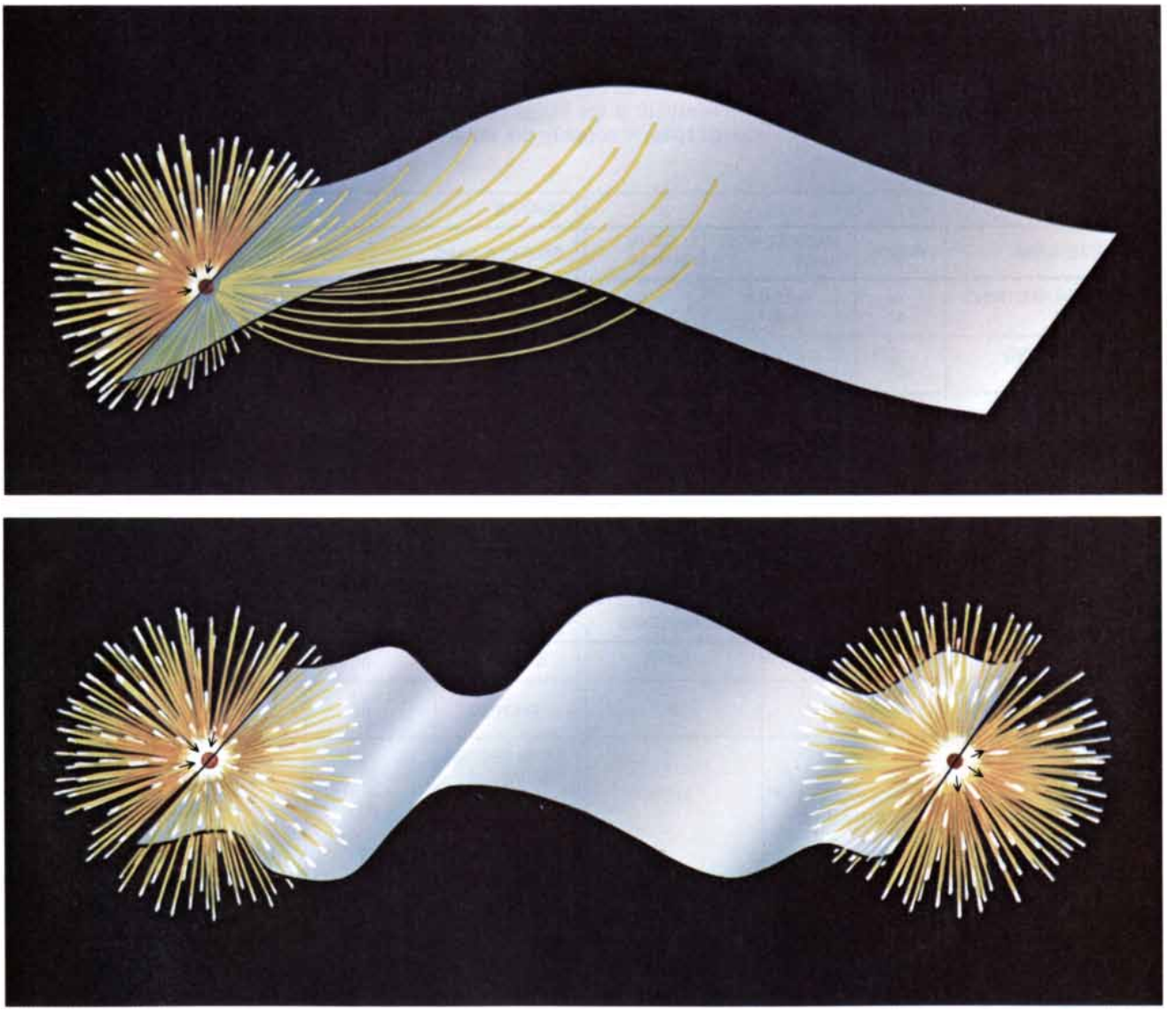
field would be a scalar field, which is one of two kinds of field important in describing the interactions of particles. A scalar field is a field in which each point has associated with it a single magnitude, or number. The other important field is a vector field: a field where at each point a vector, or arrow, is drawn. A vector has both a magnitude, which is represented by the length of the arrow, and a direction. The electromagnetic, weak and strong fields are all vector fields. (The gravitational field is a special entity called a tensor field.)

The proposed Higgs field must be a scalar field, because if it were a vector field, the mass of a particle would in general depend on the particle's alignment with the field. Stated in a somewhat oversimplified way, the mass of a person would change if he or she turned around while standing in the same place. In other words, the Higgs field is "spinless."

Because the Higgs field is spinless, the Higgs boson must also be spinless. Spin, as applied to elementary particles, is a quantum-mechanical property roughly equivalent to the classical spin of a rotating ball. Elementary particles can take on only integer (0, 1, 2 and so on) and half-integer (1/2, 3/2 and so on) values of spin. Particles that have integral spin are called bosons and particles that have half-integral spin are called fermions. Bosons and fermions have sharply differing properties, but I shall not delve into that topic here.

The Higgs boson is called a scalar boson because it has a spin of 0. Most other bosons associated with fields are thought to be vector bosons: particles that have a spin of 1. The photon, gluon and W^+, W^- and Z^0 particles, for instance, are spin-1 bosons.

Since vector bosons are typically associated with the fundamental forces of nature and the Higgs boson is a scalar boson, the force by which particles couple to the Higgs field must be a new force. It is introduced explicitly and solely as a mechanism to improve the mathematical consistency of the standard model. The Higgs force behaves mathematically in a similar manner to the recently publicized "fifth force" reported by Ephraim Fischbach of Purdue University. The



MAGNETIC MONOPOLES should exist if the Higgs boson exists. Classically, of course, magnetic monopoles are not found because when a bar magnet is cut in half, two smaller bar magnets are created—not isolated "north" and "south" poles. Magnetic monopoles could, however, be formed by sweeping magnetic field lines under the Higgs "rug" (*top*). The bottom illustration shows a pair of monopoles. Although there have been scattered reports of finding monopoles, none of them has been substantiated to date.

proposed Higgs force is, however, weaker and has a much shorter range than the "fifth force."

The Higgs force is not a universal force, because it couples differently to different particles. Specifically, if a particle is observed to have mass, the strength of the coupling to the Higgs field is assumed to be whatever quantity is necessary to generate precisely that mass. Presumably the Higgs field does not couple to the photon, since experiment shows the photon is massless. But apparently it couples to the W^+, W^- and Z^0 particles, because they do have mass. It should perhaps be noted that particles could have a mass of their own, in addition to what they are thought to acquire from the Higgs field. Curiously, however, in the standard model not a single particle could have a mass of its own without destroying the mathematical completeness of the theory.

From a physical point of view little is gained by proposing that the Higgs boson accounts for mass. It is not known, for example, why the Higgs field should couple more strongly to some particles than it does to others. Nor do investigators understand how the mass of the Higgs boson itself (which is not known) comes about, although it is generally presumed to be dominantly through a self-interaction with the Higgs field. In this sense ignorance about the origin of particle masses is replaced by ignorance about particle-Higgs couplings, and no real knowledge is gained.

Moreover, the introduction of the Higgs boson creates a significant problem with respect to the "holy" field of gravitation. The equivalence of mass and energy implies that the graviton, which couples to anything that carries mass, should couple to anything that carries energy, including the Higgs field. The coupling of the graviton to the Higgs field—ever present in all space—would generate a huge "cosmological constant": it would curve the universe into an object roughly the size of a football. If the Higgs boson is assumed to have roughly the same mass as the weak vector bosons, the energy density of the Higgs field in the vacuum would be 10 trillion times greater than the density of matter in an atomic nucleus. If the earth were compressed to this density, its volume would be approximately 500 cubic centimeters, or a bit more than the size of a soft-drink can. Needless to say, this is contrary to experiment.

The theorists' way out is really something. It is assumed that the "true" vacuum (one without a Higgs field) is curved in a negative sense: it has a cosmological constant equal in magnitude but opposite in sign to the one generated by the Higgs field. The introduction of the Higgs field then flattens out space to make precisely the universe as we know it. This solution is, of course, not very satisfactory, and many ingenious attempts have been made to solve the problem of the huge cosmological constant. None of the attempts has succeeded. If anything, matters have grown worse because theorists keep dumping more particles and fields into the vacuum.

LEPTONS

PARTICLE NAME	SYMBOL	MASS AT REST (MeV)	ELECTRIC CHARGE
ELECTRON NEUTRINO	ν_e	ABOUT 0	0
ELECTRON	e^-	0.511	–1
MUON NEUTRINO	ν_μ	ABOUT 0	0
MUON	μ^-	106.6	–1
TAU NEUTRINO	ν_τ	LESS THAN 164	0
TAU	τ^-	1,784	–1

QUARKS

PARTICLE NAME	SYMBOL	MASS AT REST (MeV)	ELECTRIC CHARGE
UP	u	310	+ ⅔
DOWN	d	310	– ⅓
CHARM	c	1,500	+ ⅔
STRANGE	s	505	– ⅓
TOP/TRUTH	t	22,500; HYPOTHETICAL PARTICLE	+ ⅔
BOTTOM/BEAUTY	b	ABOUT 5,000	– ⅓

FORCE	RANGE	STRENGTH AT 10^{-13} CENTIMETER IN COMPARISON WITH STRONG FORCE
GRAVITY	INFINITE	10^{-38}
ELECTROMAGNETISM	INFINITE	10^{-2}
WEAK	LESS THAN 10^{-16} CENTIMETER	10^{-13}
STRONG	LESS THAN 10^{-13} CENTIMETER	1

CARRIER	MASS AT REST (GeV)	SPIN	ELECTRIC CHARGE	REMARKS
GRAVITON	0	2	0	CONJECTURED
PHOTON	0	1	0	OBSERVED DIRECTLY
WEAK SECTOR BOSONS: W^+	81	1	+1	OBSERVED DIRECTLY
W^-	81	1	–1	OBSERVED DIRECTLY
Z^0	93	1	0	OBSERVED DIRECTLY
GLUONS	0	1	0	PERMANENTLY CONFINED

STANDARD MODEL of elementary-particle physics holds that there are 12 fundamental constituents of matter (*top*) and four basic forces (*bottom*). The constituents of matter are divided into two groups of six: leptons and quarks. Leptons exist independently, whereas an individual quark has never been isolated. Quarks are always part of larger particles such as protons and neutrons; a proton, for instance, is thought to be made of two up quarks and one down quark. The particles interact with one another by means of the four forces. Each force in turn has a particle associated with it (called a boson) that conveys the force. The Higgs force, if it exists, would be a fifth one, which would be mediated by the Higgs boson. The masses of the fundamental constituents are given in millions of electron volts (MeV) and the masses of the particles that convey the forces are given in billions of electron volts (GeV).

Perhaps somehow the universe became flat from the dynamics of the big-bang explosion, which is believed to have created the universe some 15 to 20 billion years ago.

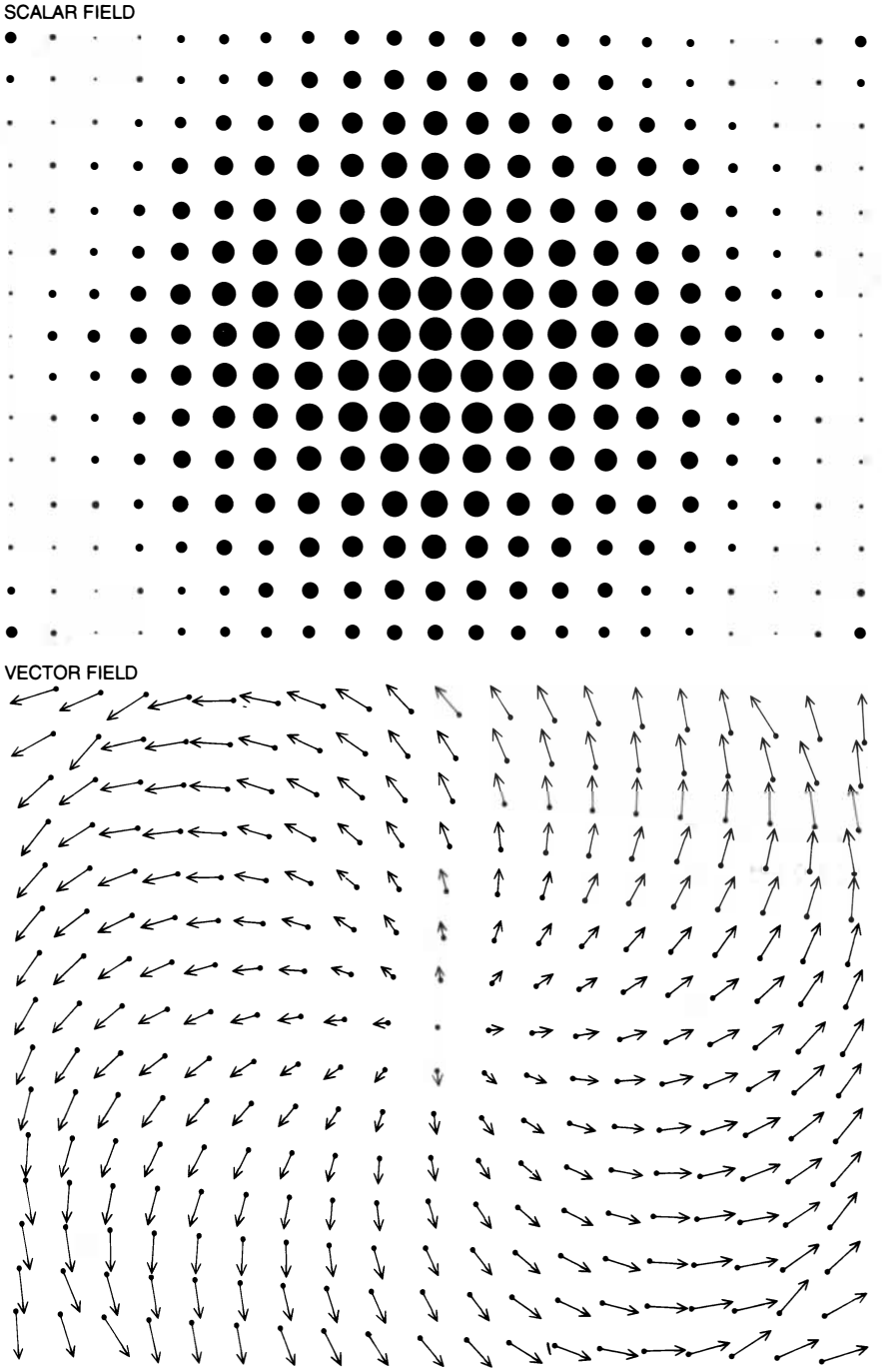
The theory as it stands, with one Higgs field, does not explicitly contradict observation, even if one must accept the incredible disappearance of the cosmological constant. Certain extensions of the theory proposed over the past decade often involve the introduction of additional Higgs fields. Although the arguments for such extensions are often compelling, the phenomena associated with these extra Higgs fields have either never been seen or contradict observed facts.

To account in an elegant way for certain symmetries observed in the strong interactions, for example, a second Higgs field was proposed by Helen R. Quinn of the Stanford Linear Accelerator Center (SLAC) and Roberto Peccei of the Deutsches Electronen-Synchrotron (DESY, the electron accelerator in Hamburg). The ensuing theory predicted a new and presumably very light particle called the axion. So far, in spite of extensive searches, the axion has not been found. In addition the theory has dramatic cosmological consequences concerning a phenomenon known as "domain walls." In general a domain wall marks where two regions of differing properties meet each other. Domain walls are, for instance, found in permanent magnets, where one region of atoms whose spins are aligned in one direction meets another region of atoms whose spins are aligned in a different direction.

It is believed that certain Higgs fields would have given rise to domain walls in the early universe. When the universe was young, the temperature was extremely hot and no Higgs field is thought to have existed. At some time the universe would have cooled sufficiently to allow a background Higgs field to come into being. Unless the cooling were completely uniform, the Higgs field would quite likely have exhibited different properties from one region in space to the next. To what extent the clash of such regions would result in visible or even violent phenomena depends on detailed properties of the Higgs fields, but one would expect some kind of clash in connection with the suggestive proposal of Quinn and Peccei.

The question is why domain walls between such regions have not been observed. It could mean that there is no Higgs field, or that nature has been careful in its use of the field. Alternatively, the walls could have disappeared early in the history of the universe. This is rather typical: one starts with an excellent argument, drags in a Higgs field and then things go wrong. It certainly inspires little faith in the mechanism altogether.

The introduction of an extra Higgs boson also creates difficulties in a model that is attracting considerable attention called the SU(5) grand unified theory. The goal of unified theories in general is to account for the four forces in terms of one fundamen-



FIELDS are important in describing the interactions of particles with one another. A field is a quantity (such as temperature) defined at every point throughout some region of space and time (such as the surface of a frying pan). Two kinds of fields are scalar fields (*top*) and vector fields (*bottom*). A scalar field is a field in which every point has associated with it a single magnitude, or number, represented here by the area of the dots. A vector field has both a magnitude, which is represented here by the length of an arrow, and a direction, which is represented by the orientation of the arrowhead in space. The electromagnetic, weak and strong fields are examples of vector fields; the Higgs field, if it exists, would be a scalar field. (The gravity field is a special entity called a tensor field.)

tal force. A step toward achieving that goal was reached over the past two decades with the introduction and verification of the so-called electroweak theory. The theory holds that the electromagnetic force and the weak force are manifestations of the same underlying force: the electroweak force. The electroweak theory was dramatically confirmed in 1983 at CERN, the European laboratory for particle physics, with the detection of the W^+, W^- and Z^0 particles.

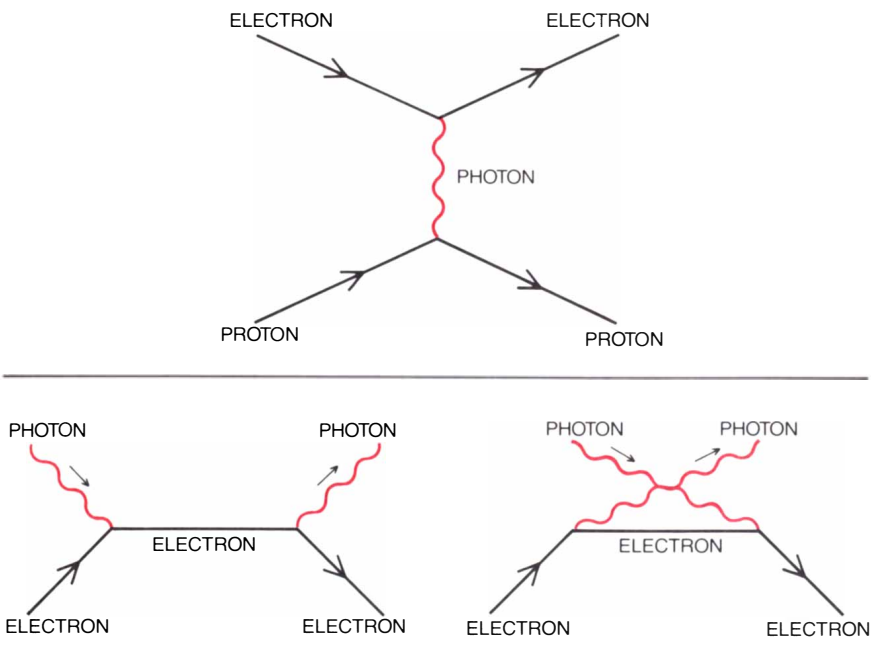
The SU(5) grand unified theory seeks to bind the strong force and the electroweak force into one common force; the designation SU(5) refers to the mathematical group of symmetries on which the theory is based. According to SU(5) theory, the strong, weak and electromagnetic forces, which behave quite differently under ordinary circumstances, become indistinguishable when particles interact with an energy of approximately 10^{15} billion electron volts (GeV).

The unification of the strong force with the electroweak force requires the existence of an additional set of vector bosons, whose masses are expected to be several orders of magnitude greater than the masses of the weak vector bosons. Since the new vector bosons are so heavy, they essentially need a Higgs field of their own. In SU(5) theory, therefore, the vacuum contains two Higgs fields that couple with different strengths to different particles.

The most important consequence of the SU(5) theory is that quarks, through the new set of vector bosons, can change into leptons. As a result the proton—that "immortal" conglomeration of three quarks—could decay into lighter particles such as a positron (a type of lepton that can be thought of as a positively charged electron) and a particle called a pion. Given the existence of two Higgs fields, the decay rate can be computed. Experiments done in recent years have not, however, found any such decay [see "The Search for Proton Decay," by J. M. LoSecco, Frederick Reines and Daniel Sinclair; SCIENTIFIC AMERICAN, June, 1985]. It would seem that there is something wrong with the SU(5) theory or the Higgs field or both. I believe the main concepts of the SU(5) theory will survive over the long run.

Moreover, if the SU(5) grand unified theory is correct and the Higgs field does exist, magnetic monopoles should have been created in the first 10^{-35} second of the universe. An example of a magnetic monopole is an isolated pole of a bar magnet. (Classically, of course, such objects are not found, because when a bar magnet is cut in half, two smaller bar magnets are created rather than isolated "north" and "south" poles.) Proponents of the SU(5) theory differ over the internal composition of the monopole and over how many monopoles should exist; it is generally agreed that the monopole should have an enormous mass for an elementary particle, perhaps from 10^{16} to 10^{17} times the mass of the proton. Although there have been scattered reports of finding monopoles, none of the reports has been substantiated; nature seems to dislike anything involving Higgs fields. The search for monopoles continues [see "Superheavy Magnetic Monopoles," by Richard A. Carrigan, Jr., and W. Peter Trower; SCIENTIFIC AMERICAN, April, 1982].

A further smattering of evidence suggests that nature has been sparing in its use of the Higgs fields—if they have been used at all. As it happens, in the electroweak theory the employment of only the simplest type of Higgs field leads to a relation between the masses of the W bosons and the Z^0 boson. The relation is expressed in terms of a factor called the rho-parameter, which is essentially the ratio of the mass of the W bosons to the mass of the Z^0 boson. (There are correction factors that need not bother us here.) The expected value of the rho-parameter is 1; experimentally it is found to be 1.03, with an estimated error of 5 percent. If there is more than one Higgs field, the rho-parameter can take on virtually any value. Assuming that the agreement between theory and experiment is not accidental, the implication is that only one Higgs field exists.



FEYNMAN DIAGRAMS are shorthand representations of a well-defined mathematical procedure for determining the probability that one particle will scatter off another. In the top illustration an electron scatters off a proton by exchanging a photon, the carrier of the electromagnetic force. The particles can also scatter off each other by exchanging two or more photons (not shown); such exchanges are statistically less likely, so that the one-photon exchange is a good approximation of reality. A photon can also scatter off an electron. Two diagrams are necessary to approximate such an interaction (*bottom*). In this case it is hard to think of scattering in terms of a force. Instead one must think in terms of elementary processes: the photon can be absorbed or emitted by an electron. There is, however, no fundamental difference between electron-proton scattering and electron-photon scattering; one can think of both types of event as elementary processes.

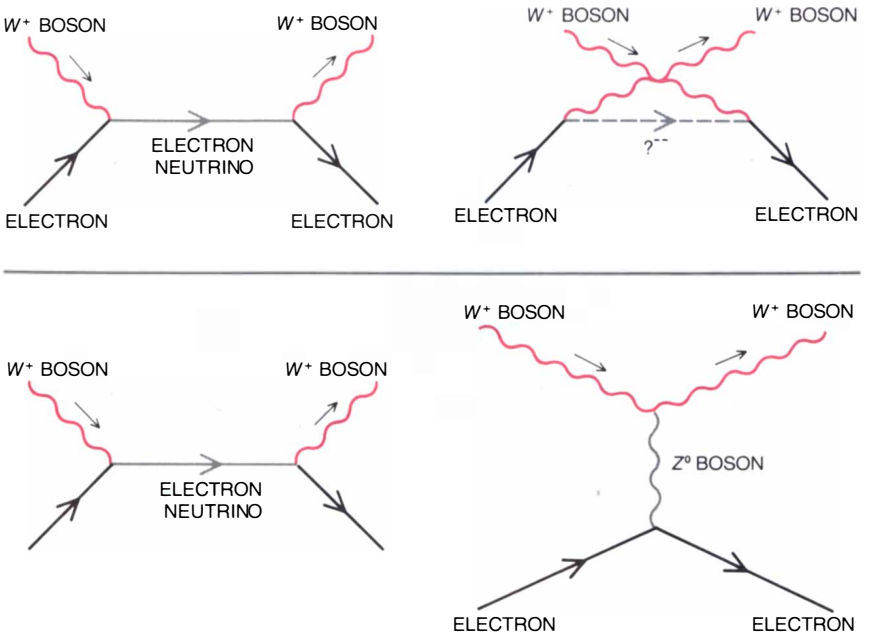
At this point it becomes necessary to question seriously whether the Higgs boson exists in nature. I mentioned above that the only legitimate reason for postulating the Higgs boson is to make the standard model mathematically consistent. Historically the introduction of the Higgs boson to give such consistency had nothing to do with its introduction to account for mass. The introduction of the Higgs boson to account for mass came out of a "model building" line, in which theories were explicitly constructed to model nature as closely as possible. Workers in this line include Sidney A. Bludman of the University of Pennsylvania, who proposed the bulk of the model containing W bosons, and Sheldon Lee Glashow of Harvard University, who incorporated electromagnetism into Bludman's model. Steven

Weinberg of the University of Texas at Austin, using methods developed by Thomas W. B. Kibble of the Imperial College of Science and Technology in London, replaced the part of the model concerning particle masses with the Higgs mechanism for generating mass. The integration of quarks into the vector-boson theory was achieved by Nicola Cabibbo and Luciano Maiani of the University of Rome, Y. Hara of the University of Tsukuba, Glashow and John Iliopoulos of the École Normale Supérieure in Paris.

All these papers were produced over a rather long period, from 1959 through 1970. In that same period many other suggested attempts at model building were also published, but none of them, including the ones I have cited, drew any attention in the physics community. In fact, most of the authors did not believe their own work either, and they did not pursue the subject any further (with the exception of Glashow and Iliopoulos). The reason for the disbelief was obvious: no one could compute anything. The methods and mathematics known at the time led to nonsensical answers. There was no way to predict experimental results.

While I was considering the body of available evidence in 1968, I decided that Yang-Mills theories (a general class of theories of which the standard model is a specific example) were relevant in understanding weak interactions and that no progress could be made unless the mathematical difficulties were resolved. I therefore started to work on what I call the "mathematical theory" line, in which little attention is paid to the extent theory corresponds to experimental observations. One focuses instead on mathematical content. In this line I was by no means the first investigator. It was started by C. N. Yang and Robert L. Mills of the Brookhaven National Laboratory. Richard Feynman of the California Institute of Technology, L. Faddeev of the University of Leningrad, Bryce S. DeWitt of the University of North Carolina and Stanley Mandelstam of the University of California at Berkeley had already made considerable inroads in this very difficult subject.

I did not finish the work either. The concluding publication was the 1971 thesis of my former student Gerard 't Hooft, who was then at the University of Utrecht. In that period few researchers believed in the subject. More than once I was told politely or not so politely that I was, in the words of Sidney R. Coleman of Harvard University, "sweeping an odd corner of weak interactions." A noted exception was a Russian group, led by E. S. Fradkin of the University of Moscow, that made substantial contributions.



WEAK VECTOR BOSON, the carrier of the weak force, can scatter off an electron in a way somewhat analogous to the scattering of a photon off an electron (*see bottom of illustration on opposite page*). The interaction at the top right, although mathematically desirable, would require the existence of a doubly charged negative particle. No such particle is known to exist. The problem is resolved by introducing a neutrally charged particle, called the Z^0 boson (*bottom right*). The existence of the boson has been verified.

Interestingly enough, the model-building line and the mathematical-theory line proceeded simultaneously for many years with little overlap. I confess that up to 1971 I knew nothing about the introduction of the Higgs boson in the model-building line. For that matter neither did 't Hooft. At one point, in fact, I distinctly remember saying to him that I thought his work had something to do with the Goldstone theorem (a concept that came out of the model-building line). Since neither of us knew the theorem, we stared blankly at each other for a few minutes and then decided not to worry about it. Once again progress arose from "Don't know how," a phrase coined by Weisskopf.

Progress in the mathematical-theory line would ultimately show that the electroweak theory becomes better-behaved mathematically and has more predictive power when the Higgs boson is incorporated into it. Specifically, the Higgs boson makes the theory renormalizable: given a few parameters, one can in principle calculate experimentally observable quantities to any desired precision. A nonrenormalizable theory, in contrast, has no predictive power beyond a certain limit: the theory is incomplete and the solutions to certain problems are nonsense.

I must point out, however, that the electroweak theory can make powerful predictions even without the Higgs boson. The predictions concern the forces among elementary particles. Those forces are investigated in high-energy-physics laboratories by means of scattering experiments. In such experiments beams of high-energy particles are directed at a "target" particle. A beam of electrons might, for instance, be scattered off a proton. By analyzing the scattering pattern of the incident particles, knowledge of the forces can be gleaned.

The electroweak theory successfully predicts the scattering pattern when electrons interact with protons. It also successfully predicts the interactions of electrons with photons, with W bosons and with particles called neutrinos. The theory runs into trouble, however, when it tries to predict the interaction of W bosons with one another. In particular, the theory indicates that at sufficiently high energies the probability of scattering one W boson off another W boson is greater than 1. Such a result is clearly nonsense. The statement is analogous to saying that even if a dart thrower is aiming in the opposite direction from a target, he or she will still score a bull's-eye.

It is here that the Higgs boson enters as a savior. The Higgs boson couples

EAST KILBRIDE, DEEP IN THE HEART OF SCOTLAND'S SILICON GLEN HAS ATTRACTED MORE U.S. INVESTMENT THAN ANY OTHER SCOTTISH LOCATION. SO SUCCESSFUL ARE THE U.S. COMPANIES ALREADY HERE THEY NOW EMPLOY NO LESS THAN 20% OF THE TOWN'S TOTAL WORKFORCE, EARNING MILLIONS OF DOLLARS EVERY YEAR.

In 1959, Kraft Foods became the first U.S. company to choose East Kilbride as an overseas location.

Since then, the town has continued to attract more and more companies from the States.

East Kilbride sits in the very heart of Scotland's Silicon Glen.

Within a 50 mile radius there are over 200 electronics companies forming not only a close-knit nucleus of suppliers, but a considerable market too.

As well as being the ideal gateway to the expanding markets in the U.K., East Kilbride is also well positioned for companies wishing to exploit the vast European markets.

Using East Kilbride as a tariff-free springboard, your company has another 600 million potential European customers.

Our labour record is excellent, with 95% of facilities enjoying a strike-free period of five years.

Because 80% of companies are union free, East Kilbride's position as a right to work town is further consolidated.

The town has a proven track record in successfully catering for industrial investment and our range of financial incentives is among the best offered by any area in the U.K.

If your company is considering an offshore project, contact Rick Packer on (617) 431 7474 at PRTM, 36 Washington Street, Wellesley Hills, Massachusetts 02181.

And join the other U.S. companies currently flying the flag in East Kilbride.

Advanced communications need advanced support. New York Telephone offers you both.

As a New York Telephone customer you have our experience and technology to support you.

We're leading the way to bring you the latest communications technology.

But communications can only be as good as the company that stands behind them.

And no one else has the resources to keep your communications running as smoothly as New York Telephone.

Our service people have an average of 15 years experience each. They monitor the network 24 hours a day and use advanced computer technology to respond instantly to your calls about service or billing.

And our computer systems are among the world's largest. They enable us to process efficiently the 21,000 service orders New Yorkers call in daily. This helps hold costs down for everyone.

If you need advanced business communications, then you also need a company that can stand behind them.

For innovative communications backed by the most experienced support team, call on New York Telephone at **1 800 942-1212, ext. 25.**

New York works on New York Telephone.

New York Telephone

A NYNEX Company

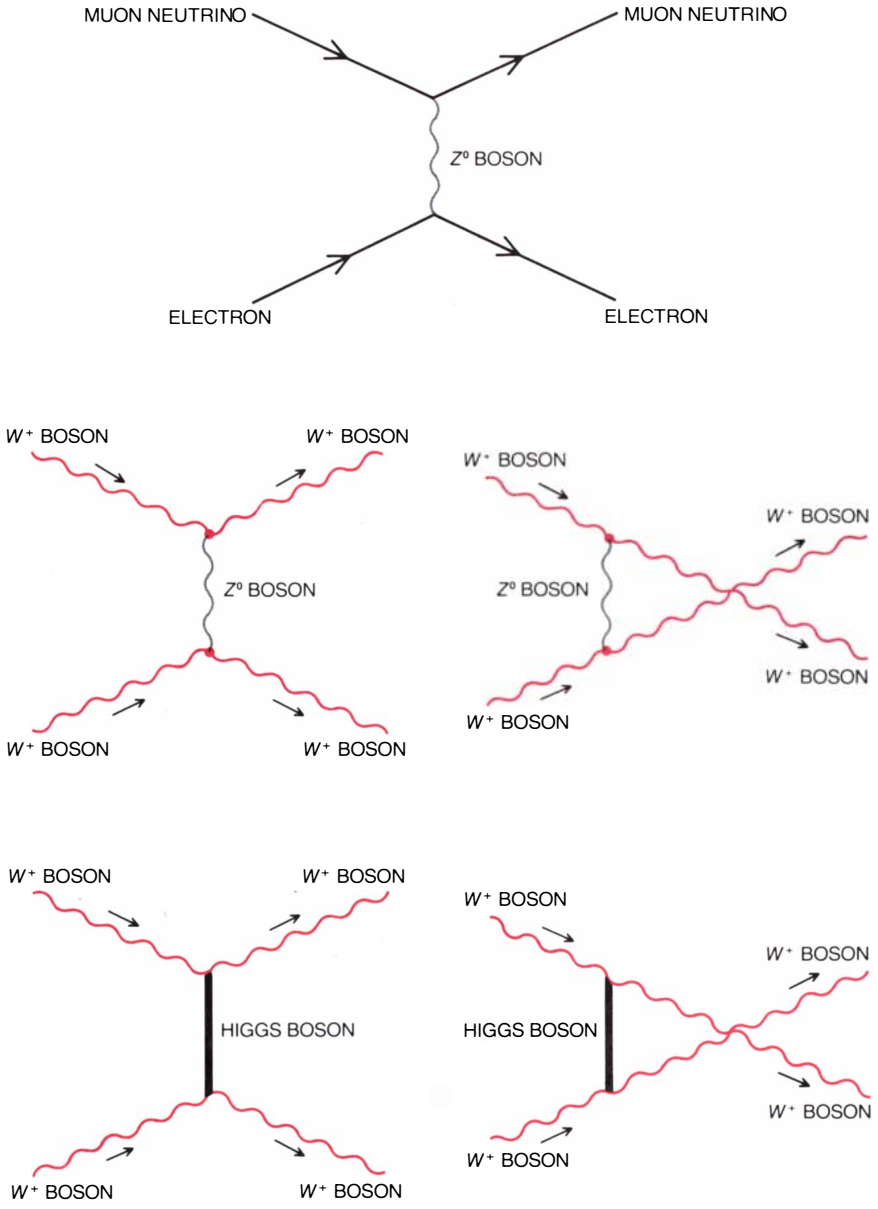
with the W bosons in such a way that the probability of scattering falls within allowable bounds: a certain fixed value between 0 and 1. In other words, incorporating the Higgs boson in the electroweak theory "subtracts off" the bad behavior. A more thorough description of the way in which the Higgs boson makes the electroweak theory renormalizable requires a special notation known as Feynman diagrams [*see series of illustrations beginning on page 80 and ending below*].

Armed with the insight that the Higgs boson is necessary to make the electroweak theory renormalizable, it is easy to see how the search for the elusive particle should proceed: weak vector bosons must be scattered off one another at extremely high energies, at or above one trillion electron volts (TeV). The necessary energies could be achieved at the proposed 20-TeV Superconducting Supercollider (SSC), which is currently under consideration in the U.S. [see "The Superconducting Supercollider," by J. David Jackson, Maury Tigner and Stanley Wojcicki; SCIENTIFIC AMERICAN, March]. If the pattern of the scattered particles follows the predictions of the renormalized electroweak theory, then there must be a compensating force, for which the Higgs boson would be the obvious candidate. If the pattern does not follow the prediction, then the weak vector bosons would most likely be interacting through a strong force, and an entire new area of physics would be opened up.

A difficulty in searching for the Higgs boson is that its mass is virtually unconstrained. As determined by experiment, the mass must be greater than about 5 GeV. Theory presents no clue as to how heavy the Higgs boson could be, except the particle would generate some of the same difficulties it has been designed to solve if its mass were 1 TeV, which is approximately 1,000 times the mass of the proton. At that point theory suggests the weak vector bosons could no longer be viewed as elementary particles; they could be composite structures made of smaller particles.

The notion of a composite structure is, of course, nothing new in the history of physics. At the beginning of the article I mentioned five known layers of structure: molecules, atoms, nuclei, nucleons (protons and neutrons) and quarks and leptons.

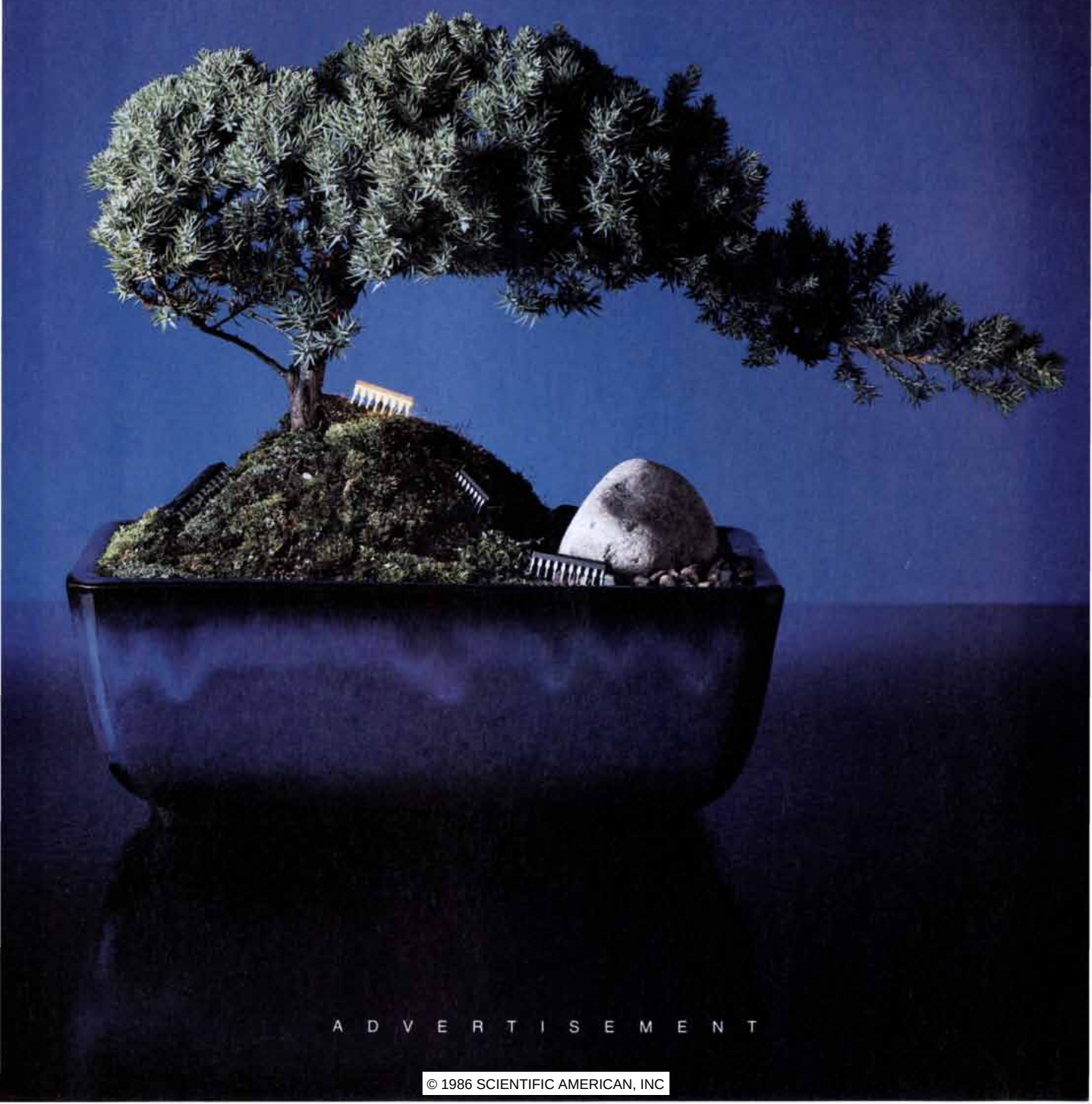
In considering the Higgs boson as a composite structure it is only a small step to suppose such "fundamental" particles as quarks and leptons are really composite structures made from still smaller particles [see "The Structure of Quarks and Leptons," by Haim Harari; SCIENTIFIC AMERICAN, April, 1983]. In a sense the notion of a sixth layer of structure, one beyond quarks and leptons, brings me full circle. Traditionally the way to account for free parameters has been to go to a deeper layer of structure. The success of composite models in predicting energy levels of atoms and nuclei suggests that mass could also be predicted by going to a deeper layer of structure. The fact that in the standard model the Higgs boson is responsible for all observed masses implies that, even if in the end there is no such thing as a Higgs boson, there is at least a common source for all masses. Searching for the Higgs boson could ultimately be the same as searching for a deeper structure of elementary particles.



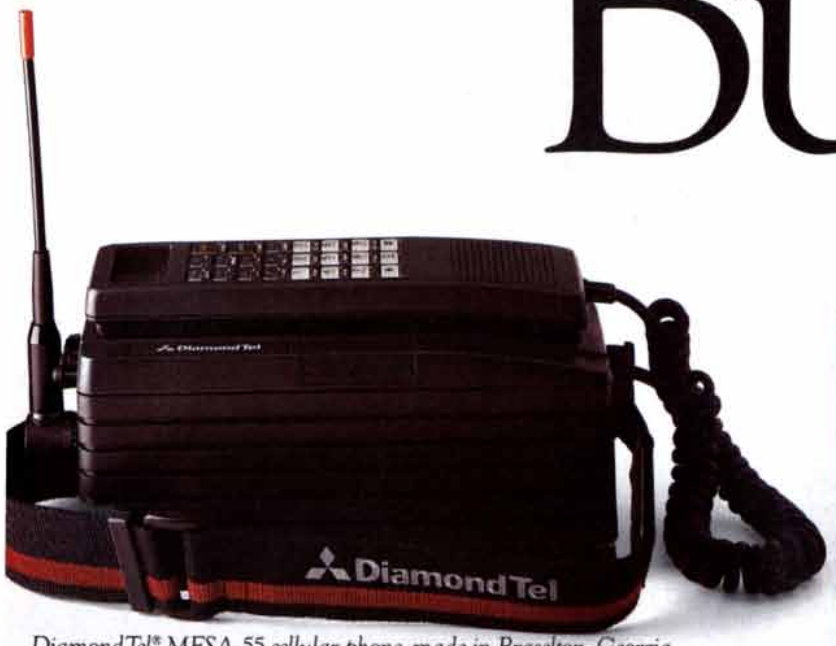
RENORMALIZED ELECTROWEAK THEORY requires the existence of the Higgs boson. A renormalized theory is one that, given a few parameters, can be applied to calculate experimentally observable quantities to any desired precision. A nonrenormalizable theory, in contrast, has no predictive power beyond a certain limit: the theory is incomplete, and the solutions to certain problems are nonsense. Without the Higgs boson the electroweak theory successfully accounts for the scattering of neutrinos off electrons (*top*). The theory runs into trouble, however, when it tries to predict the interaction of W bosons with one another (*middle*). Specifically, the theory indicates that at energies above one trillion electron volts (TeV) the probability of scattering one W boson off another W boson is greater than 1. Such a result is clearly nonsense. The theory is renormalized by introducing the Higgs boson (*bottom*). Plausible predictions can be realized by effectively "subtracting" the set of illustrations at the bottom from the set in the middle.

JAPANESE TECHNOLOGY TODAY

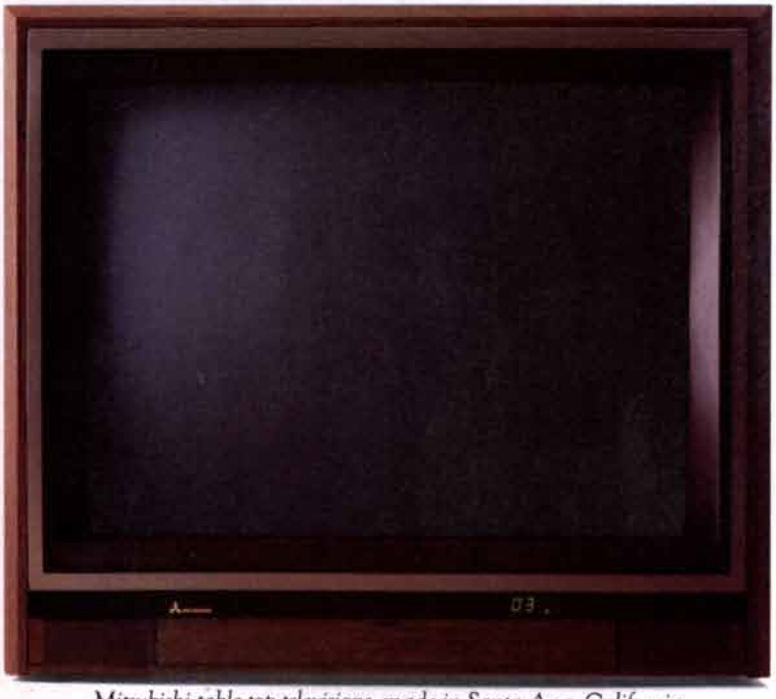
The legendary Japanese bonsai, meticulously nurtured and shaped, is a perfect miniature of a tree thousands of times its size. Today's Japanese technology represents the same tradition of craftsmanship, now extended beyond Japan and throughout the world. Microcomputer and other instrument components are just a part of the Japanese contribution to world technology, as described in the following report from inside Japan.



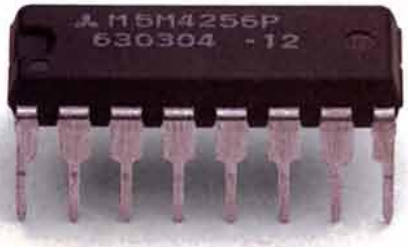
Buy An



DiamondTel® MESA 55 cellular phone, made in Braselton, Georgia.



Mitsubishi table top televisions, made in Santa Ana, California.



Mitsubishi semiconductors, made in Durham, North Carolina.

Mitsubishi big screen televisions, made in Santa Ana, California.

nerican.

Luma™ video phone, made in Milpitas, California.

With a name like Mitsubishi, it's a little difficult to get people to think of us as anything other than a Japanese company.

But it doesn't stop us from trying.

Consider the products pictured here, for example. Each one was designed and manufactured in America. By Americans. For Americans.

It's part of our on-going policy of being a good corporate neighbor in this country.

One that employs over 2,000 people. Purchases parts and materials from local suppliers. Reinvests profits back into American operations. And does its best to strike a balance between imports and exports.

As far as we're concerned, anything less would be un-American.

MITSUBISHI ELECTRIC

Japanese industry today is developing most of its own technology and investing heavily in basic research. This investment, which has been mounting since shortly after World War II, has produced significant results. Not only has Japanese industry surpassed its Western counterparts in the annual flow of patentable inventions, but the net outflow of technology has been climbing dramatically and the number of scientific papers presented by Japanese scientists at international conferences has risen sharply. By 1972, exports of technology on a current contract basis surpassed imports, and by the beginning of the present decade were running at approximately 217 percent of imports.

This shifting balance reflects at once the technological achievements of Japanese industry and the value placed on those achievements by the world marketplace. It also testifies to the willingness of Japanese industry to make this advanced technology available on terms reasonably similar to those of other advanced countries, and very often on more advantageous terms.

. . . . . .

While in the past the perceived imperative was to export to be able to afford imported raw materials and energy resources needed for economic growth, it is now vitally necessary to import to be able to export products of Japan's high-technology industries. Instead of acting as a goods supply center to the world, as business and government saw the proper role of a processing economy, adjustments must be made in institutional arrangements to transform Japan into a global-technology supply center.

For a country whose manufacturers have gained a reputation of invincibility in world markets during the 1980s and which has seen its current accounts surplus climb vigorously from US$24.2 billion in fiscal 1983 to US$51 billion last year, the sudden conversion did not come easily. A year ago, most major export products still held commanding advantages abroad. Overseas shipments of automobiles, video tape recorders, CD players, copiers, electronic typewriters, printers, floppy disks and semiconductors were gaining market share in the North American and European markets. Yet within a year all that seems to have changed.

Painfully, leading Japanese exporters have watched their market power dwindle under the impact of several major political measures, including the following:

In 1985, the EEC Commission, in an exceptionally severe move, imposed high duties on the import of Japanese electronic typewriters and followed in 1986 with high tariffs on photocopiers. U.S. semiconductor makers, though still holding the lion's share of the market at home and abroad, sought and obtained governmental support to check Japanese imports of microelectronic devices and force the Nakasone government to assure a specific minimum market share of the Japanese market to U.S. manufacturers. To cap it all, the yen was revalued, driving exchange rates upward from ¥260 per dollar to the unprecedented level of ¥150 per dollar.

Limits to Growth

The message was unmistakable, in both macro- and microeconomic terms. The days of external-dependent economic growth for Japan were ending; the carrying capacity of the global trading system had been strained to its utmost limits, providing 3.6 percent of Japan's total GNP in 1985 and accounting for a substantial share of its growth. Businessmen competing for world markets began to conclude that marketing products "Made in Japan" had limited prospects as a winning strategy.

What seemed a dramatic and crippling reversal has been accepted calmly by Japan. By the end of August signs of strain had become apparent. Profits of export industries had declined in the first half and bankruptcies of small and medium-sized export-oriented firms had risen, so no one was particularly surprised when the Economic Planning Agency announced in late Summer that signs of slump were reflected in the economic indicators.

Discounting the bad news, the Tokyo stock market soared confidently upward. Bullishness abounded. The battering taken in the foreign-policy arena did not have serious negative consequences for the majority party. With the full endorsement of the zaikai (the acronym designating the world of big business), the Liberal Democratic Party won a resounding victory at the polls, giving Prime Minister Nakasone the political clout to extend his stay in office.

And since American and European companies have relatively little at stake in the Japanese market, retaliation to protectionist measures was not really an option to be considered.

Most Japanese firms had anticipated the likelihood, if not the timing, of protectionism. Protectionist pressures on the semiconductor front had been building in the U.S.A. ever since Japanese industry took the lead in 64K DRAMs in the early 1980s. The EEC Commission had been threatening action to check Japanese imports for at least as long.

Equally important, most major Japanese exporters are horizontally diversified and vertically integrated. Higher yen value means lower-cost imports of materials, components and even the complete products of overseas factories. Since Toyota, Matsushita, Toshiba and other manufacturers, whose principal exports are preferred by foreign buyers, also provide the Japanese market with houses and other products, many of these companies and their affiliates stand to become major beneficiaries of yen revaluation.

Add to that the advantages of cheap foreign investment in overseas manufacturing, which leading international players flush with funds are capable of exploiting fully. Factory sites, plant and equipment purchased in mid-1986 cost roughly 40 percent less than a year earlier, reducing breakeven points well below those of Japanese production facilities.

All things considered, yen revaluation and the accompanying protectionist measures look very much like a formula for growth for Japanese industry. Experience suggests that calm determination can make it so.

Most important, the shock treatment provided by a higher yen serves to clarify objectives and define solu-

JUST ANOTHER LITTLE MIRACLE FROM JVC.

Two years ago, when JVC first built the smallest VHS camera/recorder ever made, people said it was a miracle. One miracle in a lifetime should satisfy most people—but not the engineers at JVC.®

THE MIRACLE OF VHS

So this year we built a camera/recorder that's not only compatible with every VHS deck in the world, but small enough to fit in the palm of your hand: The JVC Mini VideoMovie. Yet another little miracle from the people who created the VHS format, the most widely used format in the world.

The Mini VideoMovie records on regular VHS tape, wound on special small cassettes. And a carrier tray is provided for playback on any VHS recorder.

THE MIRACLE OF JVC TECHNOLOGY

Weighing in at just over three pounds, including battery and cassette, the JVC Mini VideoMovie has Instant Playback—Power Zoom—Auto Focus and other features you would expect to find only in larger, heavier camera/recorders.

The video image is read by a durable advanced CCD chip. Also, the HQ (High Quality) recording system developed by JVC allows you to record up to an hour with exceptional picture quality.

So pick up the JVC Mini VideoMovie. You'll discover why this camera/recorder was built upon the notion that at JVC miracles never cease.

JVC

Mini VideoMovie

COMPANY OF AMERICA, Consumer Video Division, 41 Slater Dr., Elmwood Park, N.J. 07407 JVC Canada Inc., Scarborough, Ont.

tions. For the past four or five years, business circles, and in particular the Japan Federation of Economic Organizations (Keidanren), have been urging government measures to end trade friction with Japan's industrial trading partners by policy that would radically alter the structure of the export-oriented Japanese economy, thus encouraging wider international technical cooperation and liberalizing product imports. It has been argued that if Japan is to play a positive role in the world free-trade system, radical adjustment of some Japanese industrial structures is essential. The mechanisms of industrial-technological change were largely in place to transform those structures and benefit from the growth that a more cooperative trade would make possible. But the political will and institutional thrust had been inadequate for their full deployment.

Now there are signs that basic issues are being addressed. The Nakasone "Import Action Program" was but a foretaste, an indicator of a more far-reaching change in public policy outlined in a remarkable document issued by the Japanese Industrial Structure Council in February this year, and then elaborated in a more comprehensive report published in April.

New MITI Plan

With the objective of drafting "The Basic Concept of Industrial Society in the 21st Century," the Ministry of International Trade and Industry's principal advisory body traced the form of a new industrial structure intended to be "in harmony with the international economic society" and contributing innovatively to the growth of that society. Noting that the world is moving from the age of Pax Americana to an era of Pax Consortia, in which order will be maintained through the collaboration of several nations, the Council stressed in its introduction that "The time has come for Japan as a nation which accounts for one-tenth of the world's GNP to change her thinking, reject national egoism, restructure her macroeconomic policies, to take measures to stabilize the exchange rate at an appropriate level and change her industrial structure, all from a global standpoint. Such changes will certainly cause pain in many fields of the domestic economy and generate friction. However, it is only by going through this pain that we shall be able to open up a new era as we approach the 21st Century."

Although this is a difficult set of objectives, Japanese policymakers and business leaders have not clearly conceptualized until now that survival and future prosperity must come from more open trading and commercial routes, the free flow of capital as well as information and a sharing of the fruits of advanced technology across boundaries.

On the premise that salvation begins at home, the Council emphasizes the importance of first taking measures to stimulate domestic consumption. Specific action to raise disposable income and increase leisure time is outlined, changes in tax and financial systems to induce housing expenditure are spelled out and massive expenditures needed for improvement of social infrastructure are prescribed. Urban renewal, industrial relocation and improved transportation, information, and communications networks are elevated to the highest priority for the central and local governments. Industry is called upon to focus research and development in microelectronics, new materials and biotechnology on these broad objectives, and to respond to needs for improved services.

The overarching theme of this vision is international, not domestic. The Council forcefully asserts that domestic policies will only have their desired effects if they are developed and carried out within the framework of close cooperative mechanisms between Japan, the United States and Europe. Noting that interdependence has increased with the free movement of goods and money, the Council makes clear its view that one country's efforts alone cannot be expected to produce sufficient results to rectify current disequilibrium among industrial nations, much less assure future stability. The Council states that in order to revitalize the world economy, securing balanced and sustained growth, the U.S. must eliminate its budget deficit, EC countries will have to restructure their industries and Japan must maintain relatively high growth rates.

Principal Challenges

To assure these high rates of growth, two principal interactive challenges are innovation and internationalization. And the burden of action lies primarily with the private sector. The government will press and encourage business to spend more on R&D, especially in basic and applications research, and to develop a new generation of frontier industries. Few government documents in any country have spelled this out quite as clearly, or in more compelling political, economic and social terms as is in the ISC's visionary map of Japan's future growth.

Internationalization is not a new theme of Japanese policy. As early as 1971, MITI began urging the internationalization of Japanese business enterprise in the face of pressures of protectionism already growing in the U.S.

With labor in chronic short supply, wages were rising rapidly, rendering cotton textiles and a broad range of labor-intensive industries unable to compete in world markets with exports of developing neighbor countries of East Asia. Since there was no way of protecting foreign markets for uncompetitive Japanese exports, the only available options were to shift production abroad, transfer technology to joint venture partners or licensees or simply abandon production entirely and shift resources to higher-value-added manufactures.

The result was a massive flow of direct investments and technology to East and Southeast Asia. By 1972, Japan had become the technology supply center of the region, and, with Asia accounting for a major share, technology exports on a current contract basis surpassed imports for the first time. Exports have continued to rise faster than imports ever since.

Behind the Japanese industrial interface with the global technological system, there has been a dramatic increase in invention and innovation. By 1968, the number of patent applications by Japanese industry exceeded that of West Germany and by 1972 it was substantially higher than the number of applications in the United States. As the level of Japanese technology advanced throughout the 1970s, both the means and the neces-

Science fiction has a way of turning into science fact.

Spaceships. Robots. For some, ideas that seem like science fiction. But for Epson, ideas like these are the seeds of the future. So when our engineers suggest an idea that seems as though it came from outer space, we give them some space to develop it.

One of those ideas was a printer small and fast enough to work in a calculator. An idea that Epson was first to realize. But we didn't stop there; we then used this expertise to develop high-performance, high-quality printers for PC's. Now Epson is the world's leading manufacturer of printers.

Another first was using liquid crystal displays for the first hand-held flat-screen color LCD television. Our experience in the original development of LCD's led our engineers to believe that a high-resolution screen could be made. So they made it, proving once again that the future is limited only by your imagination. A wristwatch TV? A TV you can hang on a wall? Who knows what's next!

Epson was also first to introduce a powerful computer small enough to fit in a briefcase. But that's hardly surprising when you consider our other small successes. Now we offer a wide selection of IBM PC compatible computers. Of course, that's to be expected. What's unexpected is more power, better graphics, and more features that are easier to use. At a value that's no fantasy.

So next time you read about an idea you think belongs in a science-fiction movie, think again. It may just be available from Epson.

Epson. Where anything is possible.

EPSON®

SEIKO EPSON CORPORATION: Suwa-shi, Japan Tel: (0266) 52-3131 Telex: 3362435 EPSON AMERICA, INC.: Los Angeles, U.S.A. Tel: (213) 373-9511/9522 Telex: 182412 EPSON DEUTSCHLAND GmbH: Düsseldorf, F.R. Germany Tel: (0211) 56030 Telex: 8584786 EPSON (U.K.) LTD.: London, U.K. Tel: (01)902-8892 Telex: 8814169 EPSON FRANCE S.A.: Paris, France Tel: (1) 47.39.67.70 Telex: 614202 TECHNOLOGY RESOURCES S.A.: Paris, France Tel: (1) 47.57.31.33 Telex 610657 EPSON AUSTRALIA PTY. LTD.: Sydney, Australia Tel: (02) 452-5222 Telex: 75052 EPSON ELECTRONICS (SINGAPORE) PTE., LTD.: Singapore Tel. 2786071 Telex: 39536 EPSON ELECTRONICS TRADING LTD.: Hong Kong Tel. 5-282555 Telex: 65542 EPSON ELECTRONICS TRADING LTD. (TAIWAN BRANCH): Taipei, Taiwan Tel. (02) 536-4339, 551-6248/9 Telex: 24444 EPSON SEGI S.p.A.: Milan, Italy Tel: (2) 670-9136 Telex: 315132 EPSON-STI, S.A.: Barcelona, Spain Tel: (3) 239-7707/8 Telex: 50129 EPSON CANADA LTD.: Willowdale, Canada Tel: (416) 495-9955 Telex: 6966633 EPSON LATINOAMERICA S.A.: Caracas, Venezuela Tel: 35-08-04 Telex: 27660 EPSON DO BRASIL INDUSTRIA E COMERCIO LTDA.: São Paulo, Brazil Tel: (011) 210-9599 Telex: 38572 EPSON CHILE S.A.: Santiago, Chile Tel: 232-46-61 Telex: 341615 EPSON ARGENTINA S.A.: Buenos Aires, Argentina Tel: 392-2620 Telex: 24353

sity of internationalization increased. By 1980, patent applications by Japanese industry had mounted to 165,000 annually, more than twice the level of the U.S. and four times that of West Germany.

Optimization of returns on the increasingly heavy outlays for R&D, as well as the necessary plant and equipment to bring new generations of products to market, required international rationalization of total operations. The same economic forces that drove the transnational development of IBM and Hewlett-Packard propelled Japanese high-technology firms into the international arena.

Japanese industrial policy had shifted from heavy and chemical industries, developed in the 1960's at high social costs in the form of urban congestion and environmental destruction, to information industries led by microelectronics, computers, communications and original consumer electronic technologies. Think tanks devoted to the study of the impact and implications of information technology abounded. All agreed: Information, the basic material to be managed, processed and distributed by the new technologies, was desirable. From the outset, the attraction for a physically resource-poor country such as Japan was that this resource can be created and can be synergetic. But in a free society, information flow is unrestrained. Information belongs to no country and cannot be contained by boundaries, territory or sovereignties. Individuals and enterprises manage information technologies optimally, on a global scale. If Japanese enterprises were to join in, they would have to develop into global enterprises.

Although never expressed in quite this way in public policy, these notions had penetrated the psyche of the Japanese technocratic community of the '70s. Literature on *Joho ka shakkai* (The Information Society) proliferated from a new school of writers: Dr. Koji Kobayashi of NEC Corporation, Hajime Karatsu of Matsushita Communications Industries, Noburo Makino of the Mitsubishi Research Institute, Professors Tadao Umesao of Kyoto University and Yujiro Hayashi of the Tokyo Institute of Technology—all men of vision, practiced in the arts and sciences of information technology, who probed the frontiers of the mind in search of new directions to a future foreseen mainly as broadly participative, knowledge-intensive and inherently international. Men of this ilk, in turn, have formulated the MITI vision of Japan's new generation of information industries.

On closer examination, there is evidence that Japan's new industrial policy describes trends that are well underway. In the past few years, about a hundred new basic research laboratories have been installed. These facilities are staffed by a new breed of young scientists and engineers, are funded by leading Japanese companies, and are sited in spacious, verdant campuses on the outskirts of Tokyo or Osaka or in the new science city of Tsukuba. Near Tsukuba there are model housing parks featuring a variety of models prefabricated or precut in factories using some of the same technologies that Japanese companies have developed in the production of ships, automobiles and skyscrapers. The notion of exchanging insularity for international integration will become a permanent feature of quality of Japanese life.

Major Japanese companies whose products have become standard household fixtures around the world, as well as many smaller components and materials makers never heard of outside of Japan, are moving production facilities closer to markets. Direct investments in overseas manufacturing rose from less than US$5 billion a year in 1980 to over US$12.2 billion in 1985, and at the rate funds have been flowing abroad in recent months, the increase in 1986 is likely to surpass the 20-25 percent rates of the first half of the decade. Banks and insurance companies have also invested abroad, anticipating the further liberalization of the Japanese financial system and responding to business opportunities implicit in the spread of Japanese manufacturing operations. Though none of these made the kind of headlines that announced Sumitomo Bank's acquisition for US$500 million of a 12.5 percent stake of Goldman,

JAPAN'S HIGH-TECHNOLOGY MARKETS: 2000 A.D.

	MARKET IN 2000 A.D. (¥ TRILLION; 1980 PRICES (CONSTANT)	PRODUCT FIELDS
NEW MATERIALS		
1. New materials market	5.4	1. High function, high polymer materials; fine ceramics; new metal materials; composite materials
2. Market of new materials application products	52.5	2. VLSI, solar battery, artificial teeth, synthetic membranes for new separation technology
(SUB-TOTAL)	(57.9)	
MICROELECTRONICS		
1. New products market	31.9	1. JJ element, GaAS (1 Mbit), OEIC
2. Application products market	131.3	2. Electronic equipment, such as computers, office machines, aircraft, precision machinery
(SUB-TOTAL)	(163.2)	
BIOTECHNOLOGY		
1. New products market	5.0	1. Agricultural chemicals, pharmaceuticals, biochip, biosensor
2. Application products market	1.6	2. Biocomputer
(SUB-TOTAL)	(6.6)	
TOTAL	227.7	

NOTE: The economic impact of product innovation of the three major technological innovations was calculated by using the input-output table for the year 2000, etc.
REFERENCES: New materials: "Present State and Outlook of New Materials" (Industrial Structure Council, March 1984); electronics: "Long-range Outlook of Electronics Industry" (Japan Electronic Industry Promotion Association, September 1985; biotechnology: "Impact of Biotechnology on Industrial Structure" (Japanese Association of Industrial Fermentation, August 1985)

THE WORLD'S MOST SOPHISTICATED CAMERA.

The Professional Maxxum 9000 is a historic advance. No camera system was ever so uniquely suited to the sophisticated needs of the professional and serious amateur photographer.

Maxxum is the world's fastest professional camera. With real time continuous autofocus to zero in on subjects faster than ever before. It's the only SLR that combines 1/4000th second shutter speed with an optional 5 fps motordrive. It even achieves X-synch flash speeds of 1/250th of a second!

The Professional Maxxum offers you a choice of interactive control backs, 3 superior flash units and the versatility of 15 interchangeable autofocus lenses, with more to come.

Only this Maxxum provides such a comprehensive array of the features your photography demands: spot/average metering; auto highlight/shadow biasing; AMPS with program shift; multi-mode exposure control; autofocus TTL flash control; rugged die cast aluminum body; multiple exposure capability; electronic depth of field preview; and optionally: focus priority shutter release; 100 exposure film back; wireless data transfer; auto-bracketing; multi-spot metering; and the list goes on.

No matter how professional your current photography is, you can achieve new levels of sophistication with the Professional Maxxum.

Be certain thatthe valuable Minolta U.S.A. 2-year camera/5-year lens limited warranty cards are packaged with your products. For more information, see your Minolta dealer or write: Minolta Corporation, Dept. MX-9, 101 Williams Drive, Ramsey, N.J. 07446. In Canada: Minolta Canada, Inc., Ontario.

THE PROFESSIONAL MAXXUM®
AUTOFOCUS 9000

ONLY FROM THE MIND OF MINOLTA

MINOLTA

Sachs, one of Wall Street's most prestigious and profitable investment banks, the total Japanese presence in the U.S. and Europe has contributed an essential financial network that will sustain increased manufacturing investment. Notably, in the past two years the electronic, automobile and metals industries have each invested approximately US$1 billion in overseas manufacturing, particularly in North America, marking a shift from the earlier reluctance of Japanese firms to undertake large-scale manufacturing abroad.

Manufacturers that have been producing television sets in the U.S. and Britain since the mid-1970s are moving VCR and CD-player output to new facilities alongside, or at new sites. Anticipating restrictions on semiconductor trade, major Japanese producers are now making 256K DRAMs in new plants strategically placed in Silicon Valley, Oregon, Scotland and West Germany. Sumitomo Electric has begun production of optical fibre in the U.S., U.K., West Germany and Australia. Each of the top six Japanese export-oriented auto and truck makers have become permanent features of the U.S. automotive industry. Toyota has embarked on a joint venture with GM, building cars in a plant at Fremont, California, and Mitsubishi Motors has a longstanding partnership arrangement with Chrysler.

Less publicized manufacturing operations moving abroad include those producing steel pipe, radiators, tires and instruments for automobiles. As final assemblers move their production off-shore, materials and components makers must follow or lose their customers. Not all of this investment follows the pattern of movement to the main consumer markets of North America and Europe. Some component makers find it more advantageous to supply these markets from labor- or resource-abundant developing countries, especially those of East Asia which are upgrading their production to more sophisticated electronic products and automobiles intended for world markets. China has earned an important place in the international production plans of several Japanese manufacturers.

Japanese companies which have long invested in internationalization now find that policy cost-effective. Several factors combine to explain this shift in production economics. The first is the lower breakeven point of foreign production facilities that can be had at a fraction of the capital outlay required at home. Minimum economies of scale are accordingly obtained at lower levels of output. This is a decisive factor in capital-intensive industries, and production of most electronic equipment, as well as automobiles, falls into this category.

Exporting Production

Once a Japanese manufacturer has developed a product and perfected its automated manufacture, embodying the most important know-how in the production equipment itself, the plant can function equally well in virtually any market—depending on the cost of capital, and the plant and equipment itself. Toshiya Inoue, JVC Senior Managing Director, makes the point clearly: "An automated plant, which incorporates hardware and software, becomes a total unitary system. Just build a prototype, let it learn its work, and production can take place anywhere, be it the United States or elsewhere. This is certainly true for cars, and VTR production operates in much the same way. With a unitary, fully automatic system, VTR production is possible in any part of the world."

What is left for Japan? "There is still important know-how in the system. If Japanese industry can come up with original products and the system to make them, we should be able to get along. If we run out of ideas, nothing will be left. In this respect, the home entertainment industry still has many seeds that it can sow," observes Inoue.

At ¥150 to the dollar, Inoue believes, manufacturing in Japan is no longer feasible. "Like it or not, we will have to move production overseas, even those high-tech operations that have become our specialty. In the case of JVC, VTR production has already begun in Europe, resulting in a proportionate decrease in output at domestic plants. And this Fall, production of videotapes will begin in the U.S."

For consumer electronics firms, the reversal will be painful. In the first half of 1986, VTR exports were still growing, up 15.5 percent over the previous year. Had everything gone according to plan, given the Japanese industry's strong technological lead, home manufacture of this product line would have served as a cash cow to sustain employment, diversification and investments in future generation technologies. Now lower-cost foreign labor, along with cheaper capital overseas, is forcing drastic strategic shifts, especially as cheaper, low-end, non-Japanese products proliferate.

However automated production systems may be, people are required for their operation, supervision and maintenance. Logistics and management require competent sophisticated personnel. Now labor and virtually all personnel costs in Japan are the highest in the world. Since most labor requirements in highly automated manufacture are indirect for white-collar personnel, productivity differentials have become a less important consideration in the total investment decision. No longer is overall production efficiency sensitive to these differentials, since direct labor is such a small proportion of total cost. Wages for indirect labor then become critical and the same productivity gap does not exist. Higher personnel costs in Japan have now become a compelling reason to move operations abroad, even for the most automated production. For components production, where direct labor is important, much of the advantage of higher productivity in Japan has been erased by the sharp rise of the yen.

Toyota's response to the change in production economics is representative. A decade ago, with robotization in full swing, Toyota's strategists did not anticipate the day when it would be more advantageous to produce automobiles abroad. Foreign protectionism was a threat but seemed to be either politically or economically manageable without radical alterations in the production system. Toyota's more advanced production technology would suffice.

Since the outset of voluntary restraints, Toyota has begun to make plans for investment in U.S. production, after Nissan, Honda and others have shown that it might pay off.

We soothe your way from West to East.

We delight you with the exotic.
We comfort you with the familiar.
We make speed serene. And precision graceful.
We ease your transition into our world
and then guide you wherever you want to go.
In everything we do, we seek to make you feel as at home in the Orient as we are.
On your next flight to the East, let us soothe your way.
For reservations, call your travel agent
or Japan Air Lines at 1-800-JAL-FONE.

JAL
JAPAN AIR LINES
We give you more of the Orient.

Atlanta New York Chicago Los Angeles San Francisco Seattle Honolulu Anchorage Vancouver

Hitachi's wide-ranging technologies in communication (from left to right): optical fibers, image signal processor, advanced telephone exchange system, satellite communication, and machine translation system.

COMMUNICATION

Communication is not simply sending a message...it is creating true understanding—swiftly, clearly and precisely.

"I know he's trying to tell me something: but what does he really mean?" In our world of proliferating technologies and new terminology, this kind of question is asked a lot. Here is what we are doing about it.

Hitachi's scientists and technicians' long-term goal is to break the language barrier. They are diligently at work today on an array of projects that will vastly improve the communications of tomorrow.

For example, we've made tremendous progress on a system to translate Japanese into English.

This system can be used to translate various scientific/technical papers and machinery/equipment manuals. Special "glossaries" can be developed to adapt it for fields as diverse as medicine, electronics and aeronautics. Further development could lead to automatic telephone translation or even portable verbal translators for travelers.

In addition to the machine translation system, Hitachi's research specialists are also developing advanced transmission systems that send your phone calls or business data across great distances using hair-thin optical fibers and laser beams. They are also working on other new methods of communications, such as advanced telephone exchange systems, satellite communication systems, TV conferences, and so forth.

At the root of much of this is our highly advanced computer technology: because Hitachi is producing some of the fastest, largest-capacity systems available today.

We link technology to human needs. We believe that Hitachi's advanced technologies will result in systems and products that are functionally sophisticated but easy to use. Our goal in communications—and transportation, energy and consumer electronics as well—is to build products and systems that will improve the quality of life the world around.

HITACHI

Hitachi, Ltd. Tokyo Japan

Toyota's management has announced its own "21st Vision," which aims at 40-percent overseas production by the end of the century.

For Yuasa Battery Company, a major supplier to the worldwide automotive industry, the road to internationalization began earlier but was raised to high priority after the changes in the value of the yen. In 1965, Yuasa drafted a 13-year business plan, whose main pillar was internationalization. Prior to that, Yuasa had constructed battery plants in Sri Lanka and Thailand upon insistence of local authorities. But the new plan envisaged a full-fledged drive for internationalization on Yuasa's own initiative.

Geographical proximity made Asia the logical place for the first stage of Yuasa's long-term plan. In the following years, the company established plants in Malaysia, Singapore, Bangladesh, Taiwan, South Korea and Pakistan. More than 20 years later, the last of the Asian facilities will be built in India, upon authorization from the local government. If it has taken longer than expected, the strategy of internationalization has proven itself effective. Presently, Yuasa has a 25-percent share of the Asian market, outside Japan, for automotive batteries. In 1981, Yuasa Battery (UK) was established to manufacture sealed lead batteries for office equipment. In April 1986, three years after production began in Britain, the company was the proud recipient of the Queen's Award for Export Achievement for supplying 60 percent of its output to other European countries. Encouraged by the high yen, Yuasa is doubling capacity at the British plant and is considering production of automotive batteries in the U.S. to supply its Japanese customers operating assembly plants there.

Realistic Global View

Views differ in Japanese business circles about the meaning and purport of sharply rising foreign investments by Japanese industry. Hisao Oka, Mitsubishi Electric's Managing Director for Research and Development, is skeptical. "I don't think that such ventures incorporate the conception of internationalization ... Regardless of talk about internationalization, our first priority is to survive in a competitive society, as an individual or as a company. Although we may have a global view in our minds, it does not relate to our everyday activities."

Canon's President Ryuzaburo Kaku takes a differing view. Fully a third of Canon's 34,000 employees are non-Japanese working in overseas operations, and over 30 percent of Canon's outstanding shares are held by foreign investors. Add to this overseas sales amounting to approximately 70 percent of the company's total revenues, and the net effect is to sensitize the entire organization to global realities.

In Kaku's words, "To contribute to the communities we serve, and thereby to the prosperity of regional and national economies, we locate our production, R&D and marketing operations as close to the consumer as possible. In so doing, we act on the premise that short-term gain which might be obtained through centralization of our activities at our home base must be subordinated to the longer-term material interests of all Canon members and constituencies." Although in the 1970s Canon established manufacturing operations in Taiwan, West Germany and the U.S., investments in copier and electronic-typewriter production in both the U.S. and Europe have increased sharply in response to increased protectionist pressures and the yen appreciation.

Seiko Epson's Senior Managing Director Susumu Aizawa shares Kaku's view. The responsible thing for an export-oriented Japanese company to do now is to rationalize production by market. "We could take an emergency refuge from the higher yen and move to Southeast Asia. But preferably, we will want to produce in the consuming nation--whether or not preparations have been made ... Specifically, we should be sharing our production activities. It is my belief that from now on, Japanese firms must consider placing more and more weight on offering what they have created. To do this, the development of original technologies becomes much more important. Internationalization of business, then, ultimately is a question of breeding a larger number of creative people in Japan."

Not everyone is enthusiastic about internationalist prospects. Some warn that American experience in internationalization should be heeded. U.S. multinationals moved their production centers to foreign locations, including East Asia, in search of lower costs and to take advantage of the strong dollar. Consumer electronics is a case in point. U.S. industry invested in Taiwan and Korea, shipping products to the U.S. as manufactures of the host country, adding to the imbalance of trade and losing the experience of production. "Eventually, the center of technology has also shifted to these regions, resulting in a 'hollowing-out' at home," JVC's Inoue observes. "As a result, you have a situation in the United States where—well, it would be an overstatement to say that everything has been lost—but first production know-how and then even design technology have been hollowed out. We would not want to follow in those footsteps of American industry, but this does not mean that we should hold onto production. Relocation appears inevitable."

Overseas investments will not resolve all questions of Japan's international economic growth. Public resentment to Japan's industrial presence in Southeast Asia in the 1970s is still recalled. Effects which expanding Japanese investments have on host countries differ, depending, among other things, on the level of development of related industries, the types of materials and parts available, patterns of competition and labor relations.

Japanese television and automobile manufacturers have already been confronted with claims from local competitors that they enjoy unfair market advantages. Avoidance of burdensome labor contracts and unionization is one of the most common charges. Domestic content is another problem. Since components are often designed to specifications of the user in Japan and are generally of a much higher quality, Japanese manufacturers tend to use their customary supply source unless something clearly better can be found. Although local companies also can, and often do, buy superior Japanese materials and components, that does not inure the general public or labor unions to the practice. And Japanese companies are more vulnerable to the criticism than the home-

grown competition.

There is no all-purpose cure for such problems. Each must be dealt with in terms of particular time and circumstances. Information can help, however. And government assistance in the building of a comprehensive foreign investment data base is now being contemplated to assist companies with feasibility studies and smooth the flow of information about investment plans between manufacturers of finished goods and makers of parts.

Long Range Outlook

The industry-by-industry outlook is for continued increases in direct overseas investment in assembly and fabrications. In-house division of labor internationally will run its course during the remainder of the century in the consumer electronics and automotive industries. Parts and materials suppliers to these industries will follow the same patterns, relocating abroad to supply their traditional customers wherever possible. Telecommunications equipment, computers and peripherals, office equipment, machine tools and robots are all likely to follow the same pattern. By the year 2000, total investments in manufacturing abroad are expected to reach US$135 billion at the then-current exchange rate, according to conservative estimates compiled by MITI in an October 1985 survey of Japanese companies. Investments overseas are expected to create more jobs in the host countries than the half-million expected to be lost in Japan. If preliminary estimates are correct, about 1,950,000 jobs will be generated in the U.S., Indonesia and Korea alone by increased Japanese manufacturing in those countries, constituting a net contribution to world economic growth.

Exporting Technology

All principal Japanese parties to the globalization process—business, labor, government and academia—are apparently fully committed. Whatever the price, Japanese industry is adjusting to achieve total global interpenetration, and foreign direct investments have been assigned a pivotal role. Furthermore, Japan has none of the political ambitions and postures of the kind that have often confused U.S. overseas investments with overseas politics.

While it sometimes may seem that U.S. legislators and civil servants are more concerned with the restriction of technology flow than with its promotion, Japanese policy has for the past 25 years encouraged the sale of technology abroad almost as enthusiastically as its purchase.

Japanese firms have demonstrated an unusual willingness to share technology. For Japanese managers mainly concerned with the synergetic development of total corporate resources for a forward technological thrust, this is plain common sense. They know by their own experience that technology does not remain the exclusive property of one company or country for very long. Optimization dictates that yesterday's technology be used to maximum advantage today, which may even mean exchanging it for access to markets or the more economical supply of essential materials and components. It is now seen in Japan as a condition of the optimal flow of goods and capital.

Patterns of Japanese technology flow are already changing and are likely to shift rapidly in the near future. In the past, Japanese industry imported advanced technology from the U.S. and Europe and exported textile, metal-fabrication, chemical-processing and assembly technology to developing countries. Now, consonant with the changing pattern of overseas investments, more high technology is being exported to advanced countries, often through intra-company channels.

As Seiko Epson's Aizawa affirms, "Overseas expansion ultimately means carrying technology abroad, often offering these to foreign partners and joining hands in manufacturing. If they like our products and want to produce them, we are willing to provide the technology. And if circumstances permit, we might buy the product from the overseas venture. This is the sort of work sharing that should be encouraged. Ideally, our corporate goal is to maintain, if possible, a ratio of 1 to 1 in the balance be-

JAPANESE PATENTS IN THE U.S.: THE TOP TEN

1980		1981		1982		1983		1984		1985	
Hitachi	391	Hitachi	498	Hitachi	475	Hitachi	440	Hitachi	612	Hitachi	711
Toshiba	246	Canon	269	Nissan	309	Toshiba	384	Toshiba	551	Toshiba	706
Toyota	242	Toshiba	257	Toshiba	295	Nissan	371	Canon	431	Canon	428
Canon	173	Nissan	247	Canon	289	Canon	325	Nissan	380	Fuji Film	383
Nissan	167	Fuji Film	238	Matsushita	202	Toyota	229	Toyota	318	Mitsubishi Electric	376
Sony	163	Toyota	220	Olympus	186	Matsushita Electric	214	Mitsubishi Electric	289	Nissan	337
Fuji Film	158	Olympus	168	Fuji Film	182	Olympus	186	Fuji Film	279	Honda	317
Matsushita Electric	151	Matsushita Electric	165	Toyota	161	Fuji Film	185	Matsushita Electric	237	Toyota	302
Olympus	134	Sony	147	Sony	135	Fujitsu	151	Sony	207	Matsushita Electric	251
NEC	91	Ricoh	143	Ricoh	127	Sony	149	Honda	198	Sony	

Source: Japan Patent Information Center

tween what we give and take."

For many Japanese firms, licensing has become good business. As the number of new technologies flowing from expanded R&D facilities now surpasses the ability of even the largest companies to commercialize them globally, export of the technology itself is often the best way to get the highest return on investment. As lead times become longer, R&D facilities more costly and life cycles of technology increasingly shorter, rapid globalization becomes more important. Selling technology entails less cost than marketing through extensive sales organizations, which most Japanese companies have not developed in all countries for all of their products anyway. And since R&D expenditures are rising sharply in response to competition, the number of technologies which can neither be commercialized through exports nor through direct investments is expected to climb sharply in the coming years.

In the forefront of this development is Hitachi Ltd., which always leads the field in patent applications at home and abroad, and in the U.S. ranks among the ten top applicants for new patent registrations. Since the early 1970s, Hitachi's overseas sales of technology have been mounting steadily. Dr. Tsuneyoshi Takasuna, Director & General Manager, Corporate Research & Development Coordination Department, notes: "Since liberalization of the patent laws, if we think in terms of the balance between the sale and purchase of technology in a given year, more money was made from sales (than was paid). But we still had contract obligations from earlier years. So, from the point of view of the total amounts that came in and went out, including payments of royalties on past licenses, we paid more than we received. In fiscal 1985, however, income exceeded payment in full measure and, on balance, we ended that year with a surplus."

Organic photo conductors developed by Mitsubishi Chemical Industries are the critical element of photosensitive drums used in laser printers and plain paper copiers. They offer greater sensitivity, image quality and printing durability than conventional inorganic photo conductors. But MCI went further than the development of the material itself. The firm perfected an entire production system, including the thin film technology used in applying the new raw material to seamless aluminum drums or aluminized polyester. This built on the company's extensive experience in a wide range of technologies relating to printing, copying and recording systems. Confident that it had a worldwide winner once the system was perfected, MCI adopted a clear global strategy. Wholly-owned subsidiaries or joint ventures would be established wherever possible to produce the conductors; elsewhere the technology would be licensed. The important consideration was to obtain the broadest possible use of the technology in the shortest amount of time. Once in place, this global network is also expected to assure the rapid diffusion of new technologies that MCI is developing for the production of OPC drums based on phthalocynanine pigments for use in laser beam, liquid-crystal, and LED printers.

"We are still small when compared with the German or U.S. giants," MCI Research Center Director Kazuo Yoshida notes. "It will take some time before we become a global enterprise. Although we are investing more abroad, we intend to continue commercialization of products together with foreign partners."

Development vs. Licensing

Takeda, Japan's foremost pharmaceutical company in terms of international business, began downstream processing in Mexico in 1957 and in developing countries of Asia in the 1960s. Its first joint venture in Europe was not formed until 1978 when it joined with Roussel-Uclaf to produce a range of the company's specialties. More recently, Takeda has formed joint ventures in Germany, Britain and the U.S. In 1985, the 200-year old Osaka pharmaceutical house made modern pharmaceutical history by becoming the first Japanese company to develop a drug abroad, an anti-cancer hormone developed and now produced jointly with Abbott Laboratories.

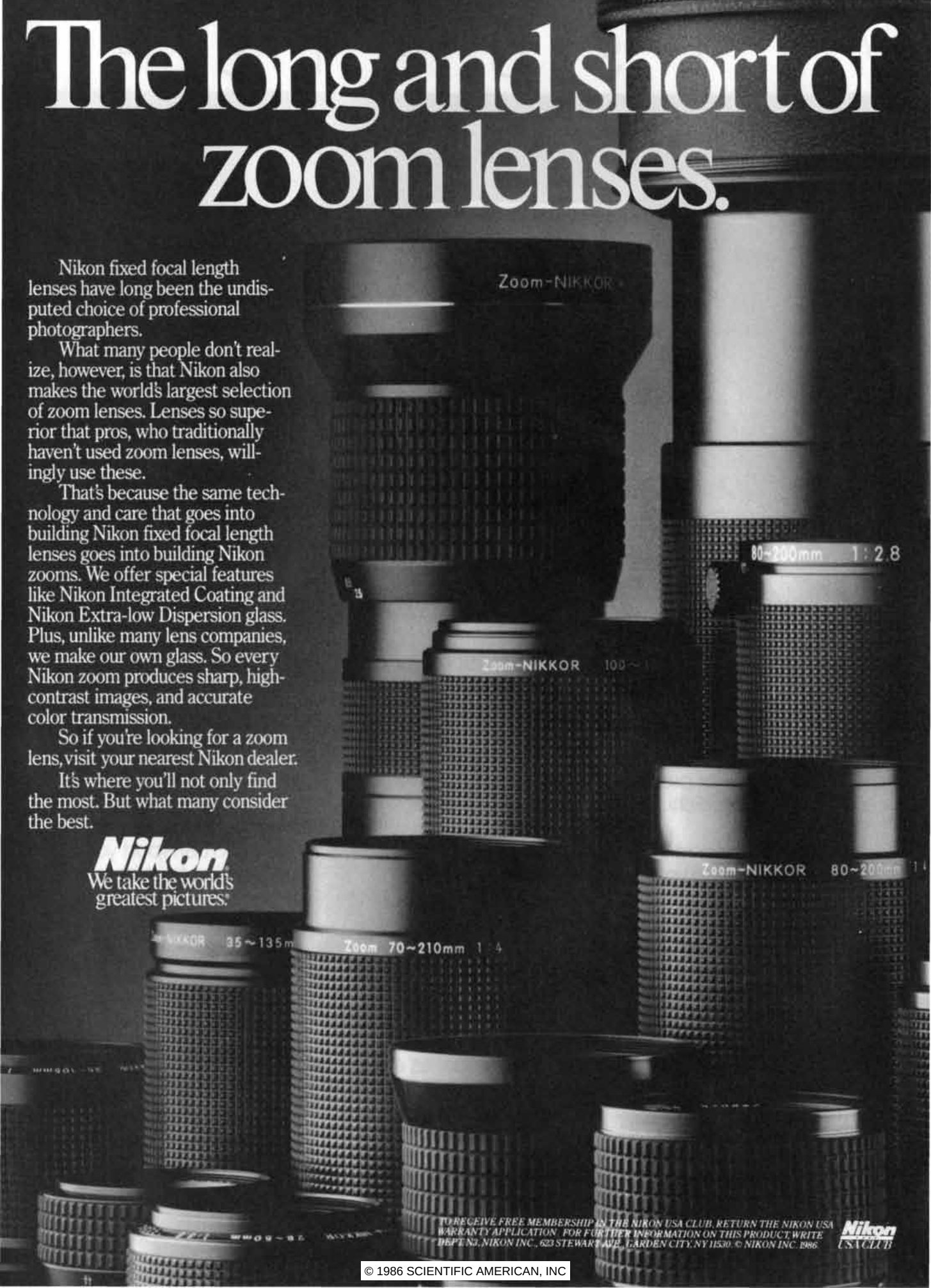
Other Japanese pharmaceutical producers are more inclined to take the licensing route. Most drug companies are far smaller than Takeda and have little international experience, but spend a larger percentage of sales on R&D. If these products are to be marketed abroad, more often than not licensing is the only option. Some leaders in new-drug development, food companies like Meiji Seika or sake brewers such as Toyo Jozo, have advanced technology but not marketing organization or experience. For them licensing abroad has compelling logic. Thus in 1983, Japan's leading whisky distiller. Suntory Ltd., sold its know how for producing gamma interferon from a synthetic gene to Schering-Plough of New Jersey. And a year later, Ajinomoto, the leading supplier of monosodium glutamate, licensed Switzerland's Hoffman-LaRoche to produce Interleukin II, an immunomodulator considered by some to have potential for treatment of cancer and acquired immune deficiency syndrome. For a growing number of companies diversifying into new fields of high technology, licensing arrangements that assure sizeable lump-sum or continuing royalty payments will become an increasingly vital aspect of their total global strategies.

For these economic reasons, technology exporting is a proposition many Japanese companies will find attractive. But for those with substantial experience in the practice, there are advantages worth more than money. Dr. Hajime Mitarai, who heads the new Canon Research Center at Atsugi, rates the cooperative relationships that can develop through licensing as its most important rewards.

Canon's license of printer technology to Hewlett-Packard some 10 years ago illustrates the point. "At the time, we licensed because we needed the money," Mitarai recalls. "It was just after the first oil crisis and business was bad. But as things turned out, a relationship that began with a one-way flow became joint development of an exciting new technology of laser-beam printing. After many meetings, confidence and trust developed, something that money cannot buy."

Mitarai continues, "Companies, after all, are people. And like the relationships between people, those between companies involve a very complex chemistry. It is not just a matter of

The long and short of zoom lenses.



Nikon fixed focal length lenses have long been the undisputed choice of professional photographers.

What many people don't realize, however, is that Nikon also makes the world's largest selection of zoom lenses. Lenses so superior that pros, who traditionally haven't used zoom lenses, willingly use these.

That's because the same technology and care that goes into building Nikon fixed focal length lenses goes into building Nikon zooms. We offer special features like Nikon Integrated Coating and Nikon Extra-low Dispersion glass. Plus, unlike many lens companies, we make our own glass. So every Nikon zoom produces sharp, high-contrast images, and accurate color transmission.

So if you're looking for a zoom lens, visit your nearest Nikon dealer.

It's where you'll not only find the most. But what many consider the best.

Nikon
We take the world's greatest pictures.

TO RECEIVE FREE MEMBERSHIP IN THE NIKON USA CLUB, RETURN THE NIKON USA WARRANTY APPLICATION. FOR FURTHER INFORMATION ON THIS PRODUCT, WRITE DEPT. N3, NIKON INC., 623 STEWART AVE., GARDEN CITY, NY 11530. © NIKON INC. 1986.

Nikon USA CLUB

MITSUBISHI

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

MONTERO

economics. Nor is nationality very important. Sometimes foreign firms are easier for us to work with than Japanese. Open minds and honesty are the essential ingredients. If the heart is small, there are very few compatible companies."

The same perception prevails at Toshiba, a company that has made international industrial cooperation an avocation. Like many major Japanese companies, Toshiba traces its roots to this kind of cooperation. In the last century, Thomas A. Edison signed in his own hand a license for the Tokyo Electric Co., precursor of the company's lighting division, to produce incandescent lamps. And as recently as the 1960s, General Electric was the largest of the company's shareholders. This relationship today takes the form of basic joint development work between Toshiba and GE on nuclear power reactors and an advanced steam turbine for improved coal-fired power generation. Other Toshiba cooperative agreements exist with:

- SGS Microelettronica SpA, a major Italian semiconductor maker, on joint development of telecommunications integrated circuits.
- Hewlett-Packard, for an exchange of researchers, including the assignment of Dr. Yoshio Nishi, a senior manager of Toshiba's Semiconductor Device Engineering Laboratory, to Palo Alto as Director of HP's Silicon VLSI Research Laboratory.
- Siemens A.G. of West Germany on a wide-ranging program of technological cooperation in semiconductors.
- Westinghouse Corp., jointly producing color display tubes for office equipment, and color picture tubes for television receivers.

The list is extensive; Toshiba has licensed widely. The Nanking (PRC) City Government has signed a technical-assistance supply contract covering production of Chinese-character printers. Joint ventures throughout Southeast Asia and other parts of the Third World are producing everything from fluorescent tubes to heavy electrical generators and data processing equipment. While the electrical giant's payments for imported technology have stabilized at about ¥8 billion a year, exports have risen rapidly.

With a number of leading companies in advanced industrial countries, a new kind of technological relationship has developed in recent years. "We have been entering into cross-licensing or balanced payment contracts with companies such as IBM, Motorola, Intel and RCA," Toshiba Executive Vice President Sakae Shimizu explains. "In such agreements, a flat rate, say ¥500 million, guarantees us the freedom to use technology we acquire in any way we choose. Cross-licensing contract provisions enable us to keep payments for technology imports constant."

The larger the company, the greater the need and capacity for managing this kind of international R&D cooperation. "The high-tech reality," Canon's Mitarai says, "is that we can do very little alone. The higher the technology, the more elusive is monopoly. No one, not even the giants, can be dominant. Technologies cannot stand alone today. Synthesis is imperative. And no one has all the pieces."

"International People"

No one believes that all the pieces will be found in Japan. Synthesis and synergy must be sought on a global basis. Manufacturing, process and design technology must be developed for application and use in countries throughout the world. "The bottom line is," says Mitsubishi Motors Executive Vice President Shinji Seki, "we must become international people. We will not be able to achieve much unless we work in an atmosphere of international cooperation. Whether we want to develop a product, conduct an enterprise, or sell a product, since our own market is not so large, we must look to international cooperation for completing the project."

Using just this kind of logic, NEC developed its entire mainframe computer line, the ACOS series, which is strong runner-up in the Japanese market and doing well abroad. NEC's ACOS 6 was originally conceived by Honeywell, with operating system and hardware designed according to needs by NEC. The specific features of ACOS 4 were prescribed by CII-Honeywell-Bull to meet particular French market needs; NEC designed the hardware and operating software to match. Inspiration for the ACOS 2 came from Honeywell, Italia; using Italian technical input, NEC enhanced the diversity of the line. The System 1000, the high-speed computer first developed by NEC in 1980, began as GE approached NEC with specifications it required. NEC had strong and enduring links with Honeywell, and NEC explained GE's interest to Honeywell. NEC designed and produced the machine in cooperation with Honeywell engineers, and Honeywell ended up supplying GE.

NEC, in this instance, serves as the keystone of a functioning worldwide system. Ownership is not a factor; rather, the system is the product of relationships between partners who have been working together for decades. The experience of co-management, drawing on the attributes of the partners, is the binding ingredient.

"This does not mean that it's all a bed of roses," NEC's Engineering Vice President Yasukuni Kotaka cautions. "Facts are, it's very hard work. Everyone involved must be sensitized to the problems and concerns of other partners. Each has different markets, therefore different strategies. Each company has it own experience and corporate culture. Japanese, European and American companies have different access to and costs of money and men; and the performance measures used for both often vary. All have peculiar internal political and policy environments, with long historical antecedents. And when it comes to technology, each partner has different objectives, capabilities and R&D methods."

The important factors in the formula for success, Kotaka says, are basically simple: complementarity and confidence. Cooperative arrangements must be rational. There must be a continuing and persistent quest for synergy of capabilities and resources in a constantly shifting and increasingly competitive environment.

Olympus Optical Co. find the attractiveness of cooperative research compelling for solid, down-to-earth reasons. Toshio Nakatsubo, Managing Director of the company's Corporate Research Division, makes the point candidly. "If we plan to offer a new product for the U.S. market, it makes sense to develop it there. Today, con-

ducting research overseas no longer costs as much. If we carry out research in Japan, we must first invest in facilities, which cost a lot. Money invested in research conducted in the U.S. is more efficiently utilized since facilities and administration are already in place."

A highly research-oriented company, Olympus maintains a large contingent of scientists and engineers abroad and invests heavily in their training. "We send out people abroad to the limit of our financial resources," Nakatsubo says. "Some we enroll in graduate programs at universities. Others we send to perform cooperative research with a university. We are also staffing research operations of joint ventures with foreign partners. In fact, participating in work on themes set up by American scientists seems to be the best strategy and leads to the least international friction."

University Linkages

High-technology management is inherently an international process in which companies must cooperate to be able to compete. Nor is the cooperation only between companies. As Japanese firms expand to R&D in basic science, they are establishing links with universities abroad. More companies operating at the forefront of microelectronics, new materials and biotechnology are sending promising researchers to U.S. and European universities, not only to learn about developments in those centers of advanced science, but to get to know the professors and students. Personal relationships will be a key to future cooperation.

For some American universities, these relationships, already cultivated by postwar generations of Japanese alumni, have begun bearing fruit. When MIT and Stanford set up their industrial liaison programs, Canon became a prime candidate for membership: Hajime Mitarai is a graduate of both institutions. Many of the 45 Japanese firms that have joined the MIT program pay US$30,000 a year for access to the university's research and the opportunity to participate in university projects. Toshiba rotates senior researchers at MIT on full-time assignment. In November 1982, to assist advanced voice recognition research at MIT, NEC funded a permanent chair; results of this work are all published and the professor holding the chair consults with other companies as well as NEC. At MIT alone, nine Japanese companies have funded chairs in high technology, at US$1 million each.

Mitsubishi Chemicals, which has also joined both the MIT and Stanford groups, is now contemplating financing of additional Mitsubishi chairs in the U.S. "We find the American universities eager to cooperate," Kazuo Yoshida says. "They want to plug into the Japanese research system, and we welcome that. At the same time, we have access to very important resources which we need. Both parties benefit."

The quest is for what Canon's Mitarai calls, "The Golden Mean," a synthesis that incorporates Western individuality and the Japanese skill of cooperative management. Ricoh Managing Director Takao Nawate has the same perspective and objective. "In my opinion, the basis of creativity is expertise plus individuality and the ability to think differently from others. In this respect, the West is still our senior. There are some interesting complementarities in this difference that we have tended to ignore. Between the creativity of the West and Japan's production technology or quality controls, some interesting deals can be made."

Nawate speaks from experience. In 1979, Ricoh established its first overseas R&D venture, Ricoh Systems, Inc. (RSI) in San Jose, on the periphery of Silicon Valley. Most of its 44 employees are Americans, but the melting-pot culture in California provides fertile ground for creativity—all focused on the development of copiers, facsimiles and semiconductors.

The number of Western firms interested in international technology-sharing is growing steadily.

Western Ventures in Japan

Multinationals once reluctant to form joint ventures in Japan now find it advantageous to team up with Japanese firms in developing and producing for their markets at home and abroad. Both U.S. and European biotechnology firms have garnered development funds through licensing or joint research with established Japanese chemical, pharmaceutical or food companies. Genex, a Maryland-based venture business, developed new biological products and had ample R&D capability for pushing forward in biotechnology, but did not have resources to expand for production and marketing. Japanese companies have been the main source of funding and have provided channels for commercialization. For example:

- Yamanouchi Pharmaceutical is manufacturing and marketing a Genex product that dissolves fibrin, paying a royalty of eight percent on worldwide sales.
- Yoshitomi Pharmaceutical Industries has contracted Genex for the development of genetically modified

TRENDS IN TECHNOLOGY EXCHANGE (TECHNOLOGY TRADE)

	TECHNOLOGY EXPORTS			TECHNOLOGY IMPORTS			
	NO. OF CASES	RECEIPTS (¥100 MILLION)	GROWTH OVER PREVIOUS YR. (%)	NO. OF CASES	RECEIPTS (¥100 MILLION)	GROWTH OVER PREVIOUS YR. (%)	PAYMENTS/ RECEIPTS (TIMES)
FY 1979	3,667	1,331	9.1%	7,012	2,410	25.5%	1.81
1980	4,103	1,596	19.9	7,248	2,395	−0.6	1.50
1981	4,877	1,751	9.7	7,207	2,596	8.4	1.48
1982	4,738	1,849	5.6	6,936	2,826	8.9	1.53
1983	6,403	2,409	30.3	7,839	2,793	−1.2	1.16
1984	5,426	2,775	15.2	7,316	2,814	0.8	1.01
FY 1979	1,087	521	10.5	1,020	268	−29.8	0.51
1980	1,237	743	42.6	919	277	3.2	0.37
1981	2,017	708	−4.7	844	249	−10.0	0.35
1982	1,970	633	−10.5	929	444	78.4	0.70
1983	2,494	749	18.3	1,073	424	−4.5	0.57
1984	1,824	909	21.4	982	318	−25.0	0.35

Source: Japanese Industrial Structures Council

organisms to produce Interleukin II.

- Green Cross, another leading Japanese drug company, has engaged Genex to develop a microbial strain that produces human serum albumin, and will market the products developed under the contract in Asia and the Western hemisphere in return for a royalty on sales.
- Mitsui Toatsu Chemicals has an exclusive license to make, use and market a microbial strain, developed by Genex under contract, that produces human urokinase.
- Mitsubishi Chemical Industries will develop and market Genex's own human serum albumin.
- Japanese companies have also been among the most avid attendants of Genex training courses in DNA technology, paying US$120,000 for a three-month session.

Typically, Japanese companies entering into contracts, joint ventures or licensing agreements provide funds for proprietary R&D; or, in the case of pharmaceuticals, do clinical evaluations and obtain regulatory approvals for marketing in Japan. Revenues received, in turn, make venture capital firms attractive when they turn to public capital markets. Biogen of Switzerland licensed or entered into joint development arrangements with at least eight Japanese firms, using contract revenue to carry it through critical early years. Genentech, which had developed annual revenues of some US$90 million through licensing and research contracts before marketing a product, derived important funding from Japan.

In the U.S. and Europe, some regard the transfer of technology to Japanese firms as a threat. Others are concerned that with rising Japanese technology exports, their own industries may become somehow dependent and lapse into subservience. In fact, both mounting Japanese technological prowess and growing technological interdependence are features of international reality. Companies like Texas Instruments and Fairchild now think of Japan as a site for melding the best of U.S. and Japanese semiconductor design production technologies; once the optimal blend is found in their production facilities in Kyushu, where both are producing their most advanced devices for world markets, the results are then rapidly diffused throughout their global production systems. To enhance this process, Texas Instruments is now building an R&D center at the Tsukuba science city, close to MITI's Electrotechnical Laboratories and 50 or so other public research institutions. TI, Fairchild, Intel and others are competing for U.S. and world markets with devices made in Japan and a great deal borrowed from Japanese technology. Now Kodak and DuPont are building new laboratories, just down the road from Mitsubishi Chemical Industries' Central Research Laboratories.

Conclusion

The three most important things to understand about current technology policy are:

- All energies are being mobilized to bring about a radical structural adjustment in industries to foster global interdependence;
- Japanese industries and government are now concentrating resources in basic research and the development of original technologies;
- Both processes will remain a largely civilian affair, for civilian purposes.

No one in Japan harbors illusions about Japanese dominance of world markets. There is, however, genuine self-confidence that by applying their full energies to the task at hand, Japanese can be creative in high technology, and that technology will be shared as Japan has shared in the technology of others.

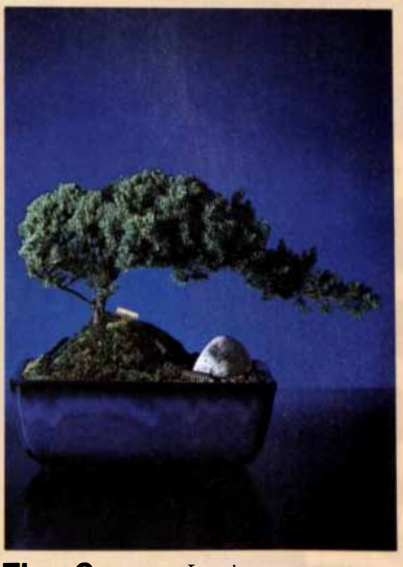
Discussion continues in Japan on the ways and means of technology policy. Questions remain about the best way to develop new international institutions to coordinate the processes of structural adjustment. Advice and input from interested partners both inside and outside Japan are sought on practices and policies. In 1987 and beyond, the best Japanese technology will in fact be truly international technology, cooperatively developed and jointly useful.

The Authors

The 1986 edition of "Japanese Technology Today" was written by Gene Adrian Gregory and Akio Etori.

Mr. Gregory is Professor of International Business in the Department of Comparative Culture at Sophia University. He is a former associate editor of *U.S. News & World Report,* a contributor to *Far East Economic Review* and author of *The Japanese Challenge in Europe.*

Mr. Etori is an award-winning and respected writer on science and technology, having covered these areas for over 20 years. He is managing editor of *Saiensu,* the Japanese-language edition of *Scientific American* and Japan's leading science publication.



The Cover: *Juniper procumbens 'nana'* bonsai tree photographed by Dan Wagner. Art direction and supervision: Edward Bell. Microelectronics: Hitachi America, Ltd.

NEW CAMRY. AMERICA'S MOST TROUBLE-FREE CAR MOVES CLOSER TO PERFECTION.

The all-new Toyota Camry is here. And perfection is within your reach.

Camry has established its reputation as the most trouble-free new car,* domestic or import, sold in America.

And the 1987 Camry is even more impressive. A new 16-valve engine gives you more highway passing power, and quicker acceleration from stop. A new aerodynamic design and a refined suspension make Camry's ride quiet and more comfortable. And front-wheel drive with power-assisted front disc brakes gives you confidence in any weather.

This Camry is designed for total customer satisfaction, and built to achieve it. So is the all-new Camry Station Wagon...even more of a good thing.

Toyota Camry. It just may be the closest a car can come to perfection.

TOYOTA CAMRY

*Based on problems encountered in first 90 days of ownership. 1985 J.D. Power & Associates New Car Initial Quality Survey.

PERFECT FOR FIVE.
Seats five adults in generous, deeply cushioned velour comfort. Bigger trunk than ever, with 15.4 cu. ft. of cargo space and a low liftover height that makes loading and unloading easy. And more, as Camry moves comfortably closer to perfection.

Get More From Life...Buckle Up!

WHO COULD ASK FOR ANYTHING MORE!

TOYOTA

In Horseheads, New York, Toshiba is taking another step in the right direction.

In 1985 Toshiba, one of the world's leading electric and electronics manufacturers, joined with the Westinghouse Electric Corporation to create a joint venture company called the Toshiba Westinghouse Electronics Corporation.

In 1986 this new company will start manufacturing color display tubes and color picture tubes in the little town of Horseheads, New York.

The tubes, which will be among the most advanced in the world, will be used not only in TV sets, but also in computer and other display terminals in offices, factories, homes and hospitals.

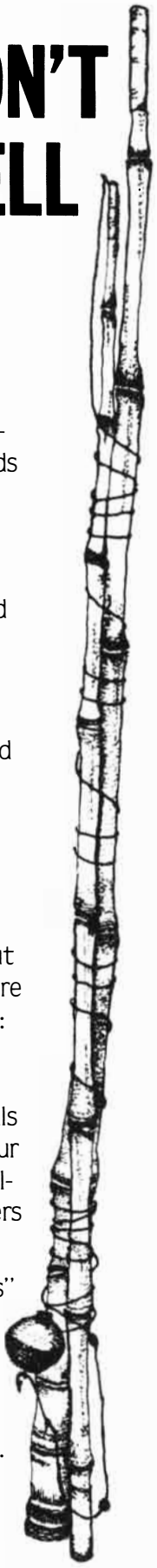
This project follows hard on the heels of other manufacturing projects Toshiba has already set up—electric motors in Houston, Texas, semiconductors in Sunnyvale, California, and color televisions and microwave ovens in Lebanon, Tennessee. It is, for Toshiba, another step in the right direction.

In Touch with Tomorrow

TOSHIBA

THE FISH WON'T BE BITING IN TWITCHELL CREEK TODAY,

OR TOMORROW, EITHER.



There was a time when with a little luck and a little patience, you could pull a fair-sized fish out of Twitchell Creek on Woods Lake in the Adirondacks.

But no more.

The fish are gone, along with the salamanders, ospreys, mayflies, swallows and myriad other creatures who once lived along the cool river banks.

They've been run off, or killed off by the rain, of all things. A deadly poisonous acid rain which has contaminated the water, choked the life out of the stream . . . and broken the delicate food chain on water and (we're finding out) on land as well.

All rain contains some acid, of course. But acid precipitation is different. And far more dangerous. Acid rain contains two killers: nitric acid and sulfuric acid which form when sulfur dioxide and nitrogen oxide mix with rain water. These two chemicals are being spewed by the ton-load into our air every day . . . emissions from the coal-burning power plants and industrial boilers our nation uses to keep going.

Fortunately, some lakes contain "buffers" . . . neutralizing agents which help lessen acid damage. But what of the others . . . in the Adirondack Mountains, in western Virginia, in the Great Smoky Mountains, throughout New England . . .

Who Will Stop The Rain?

The Izaak Walton League is working to do just that right now. The League was formed in 1922 by a handful of sportsmen who wanted to combat water pollution. And it endures today as a grassroots conservation organization composed of 50,000 fishermen, hikers, hunters and campers who speak out—and work hard—to protect wild America.

Congress passed a Clean Air Act in 1970—soon up for renewal—which set allowable limits for sulfur dioxide emissions from power plants. Some would like to relax those laws now . . . asserting the regulations will retard energy development. We disagree. We want stricter regulations to reduce these emissions still further. And we'll get them. Once and for all. Because there's something at stake here far greater than fishing. And that's life itself.

For more information on our activities, write:

THE IZAAK WALTON LEAGUE OF AMERICA

Izaak Walton League
1800 North Kent Street
Arlington, Virginia 22209

Antarctic Fishes

Most fish species perished when the Antarctic Ocean turned cold and icy, but fishes of one suborder, Notothenioidei, met the challenge. They survive by making biological antifreezes and conserving energy

by Joseph T. Eastman and Arthur L. DeVries

In February of 1899 the British ship *Southern Cross* put 10 men ashore at Cape Adare in Antarctica, thus beginning the first expedition to endure a year on the world's southernmost continent. Many zoologists credit the expedition, which ushered in the "heroic" era of Antarctic exploration, with a discovery that has intrigued them for almost a century: the coldest marine habitat in the world is actually alive with fishes. The team's zoologist, Nicholai Hanson, did not survive the year on the icy land, but before his death he collected examples of previously unknown fish species.

Almost a century later we and other investigators are still trying to understand the adaptations that enable fish to survive in a region once thought to be virtually uninhabitable. Of particular interest are the evolutionary adaptations of the suborder Notothenioidei, a group of teleosts, or advanced bony fishes, that are related to the perchlike fishes common in virtually all marine habitats. This suborder of between 90 and 100 species is primarily confined to the Antarctic region; there it dominates, accounting for an estimated two-thirds of the fish species and 90 percent of the individual fish in the area.

We have concentrated much of our study on two noteworthy adaptations. The first one, clearly crucial to cold-water survival, is the ability to produce compounds that have powerful antifreeze properties. Such compounds depress the freezing point of body fluids. The second adaptation is the development in certain species of "neutral" buoyancy, or weightlessness in water. Weightlessness spares fish from having to expend precious energy on flotation. It appears to have enabled at least two species of Antarctic notothenioids to radiate from the bottom of the sea, where most of them dwell, to the midwaters, which are underutilized.

Several geologic and oceanographic events account for the overwhelming dominance of the Antarctic notothenioids. During most of its history Antarctica was joined to the other southern continents in a great mass known as Gondwanaland. The agglomerate began to break apart perhaps 80 million years ago. The water surrounding Antarctica probably was reasonably warm in the early years, at least in places. In a recent survey of 38-million-year-old fossils found on Seymour Island, one of us (Eastman) and Lance Grande of the Field Museum of Natural History in Chicago found that the coastal waters once supported sharks, saw sharks, rays, ratfish, catfish and other temperate-water groups now absent or poorly represented in the Antarctic.

At about the time the Seymour Island fossil deposit formed, Antarctica became fully separated from Australia and the tip of South America, and its shores were surrounded by vast expanses of cold, deep ocean. Complex environmental changes contributed to the cooling of the Antarctic waters. In the ocean itself one of these changes was the formation of the Antarctic Convergence. This pattern of ocean currents, which lies between 50 and 60 degrees south latitude, surrounded the newly isolated continent and eventually became a formidable thermal barrier, impeding the inflow of warm currents and to a great extent warm-water fishes from the north.

Presumably the cold water caused temperate fishes to disappear—but not the notothenioids. They apparently began to evolve under the influence of the cold. Today, because of both the cold and certain other factors, the Antarctic Ocean has less diversity than even the Arctic Ocean, which supports one and a half times as many species and twice as many families. Among the other factors accounting for this lack of diversity are a paucity of island groups and a deep continental shelf at Antarctica's margin; the shallow water that typically surrounds islands and coastal shores is a prime habitat for many fish species elsewhere.

To learn about the nature of the adaptations that have enabled notothenioids to evolve and thrive where other fishes have failed, we and our colleagues periodically travel to Ross Island, about 400 miles south of Cape Adare. For nearly 30 years the National Science Foundation has maintained a biological research station on this small, volcanic strip, which is separated from the mainland by the 40-mile expanse of McMurdo Sound.

One of us (DeVries) began going to the island 20 years ago, primarily to clarify the mechanism by which Antarctic fishes avoid freezing. By that time certain background information had already been uncovered about the conditions in McMurdo Sound. In 1961, for example, Jack L. Littlepage, then at Stanford University, established that the average yearly temperature of McMurdo Sound is -1.87 degrees Celsius, with the range varying only between -1.4 and -2.15 degrees. In the austral summer, from December through February, temperatures rise from -1.9 to -1.8 degrees. Even in summer the water under the ice receives less than 1 percent of the sunlight that strikes the surface, but this is an improvement over the total darkness that prevails for four months of the year.

More hazardous for Antarctic fishes than the dark and the cold is the danger posed by multiple layers of ice. Some two to three meters or more of "annual" sea ice cover the water for at least 10 months of the year—until summer, when storms usually break up the ice and wash it out to sea. An additional meter or two of platelet ice (large,

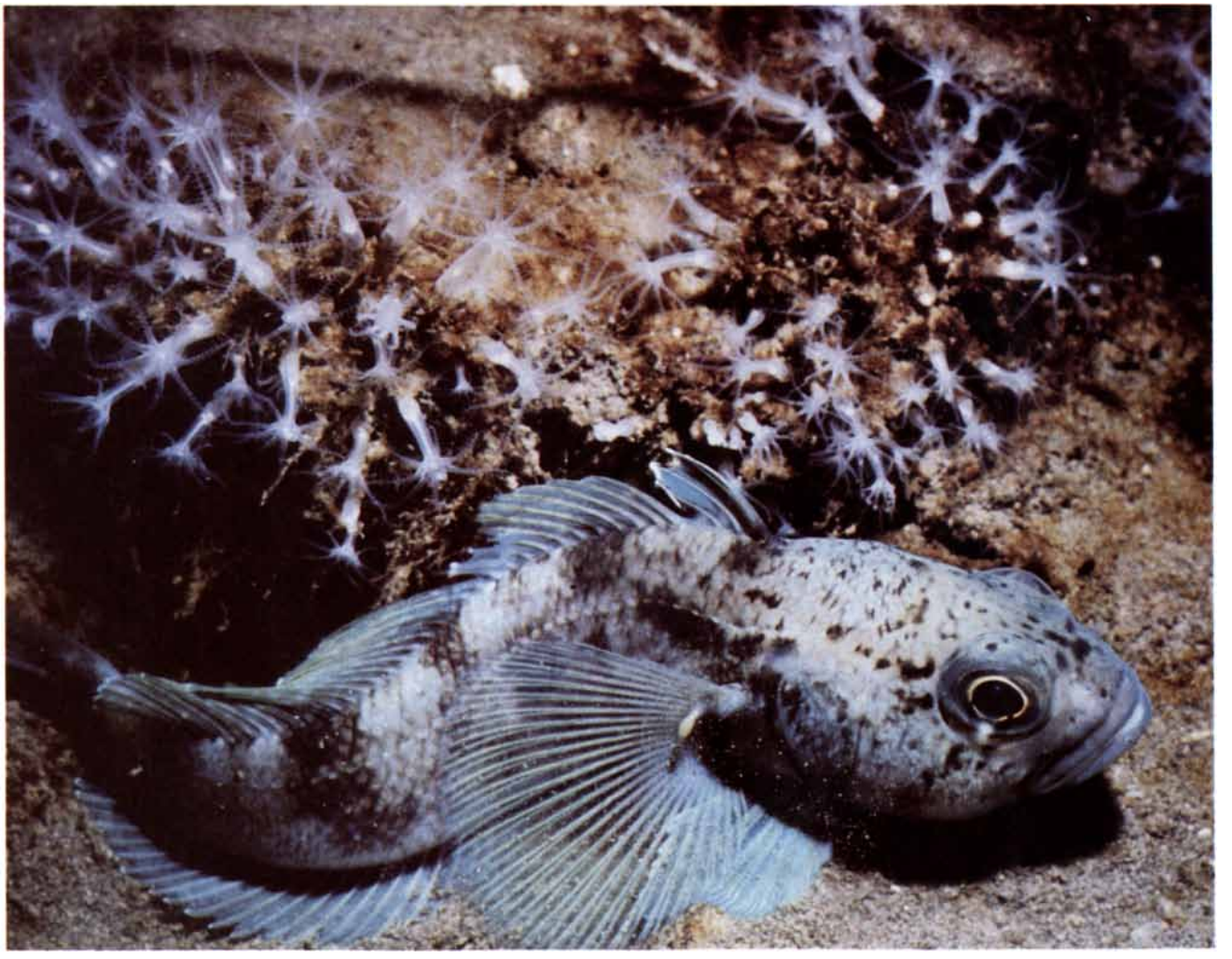
elongated, loosely aggregated crystals) adjoins the undersurface of the annual ice, disappearing starting in mid-December. Part of the year yet another layer of large crystals, called anchor ice, coats the bottom of the sound wherever the water is no deeper than 30 meters.

Ice, as one investigator demonstrated in the 1950's, is a threat because it easily penetrates a fish's gills and integument. Fish, which are cold-blooded and typically have essentially the same temperature as that of their environment, can survive when their blood cools as much as one degree below its equilibrium freezing point: in practical terms, the point at which ice crystals can form. On the other hand, fish can endure such "undercooling" and retain the liquidity of their fluids only if no ice enters the body. When ice is present around a fish that has been undercooled by as little as .1 degree the ice can rapidly propagate across the integument and freeze body fluids. In the presence of ice most tropical and temperate fishes freeze when their body fluids cool to approximately −.8 degree. McMurdo Sound notothenioids, on the other hand, freeze only when their temperature goes down to approximately −2.2 degrees.

With these data in hand, DeVries and his students set about determining the relative contributions of various substances to the freezing-point depression of McMurdo Sound fishes. For most marine fishes, salts (particularly sodium chloride) in the body fluids account for 85 percent of the depression below zero degrees, which is the standard freezing point of pure water. The remainder of the depression can be attributed to small amounts of potassium, calcium, urea, glucose and amino acids, all common constituents of blood and tissue fluids.

In the McMurdo Sound fishes, the workers found that sodium chloride and other ions and small molecules account for only from 40 to 50 percent of the freezing-point depression, even though the concentrations of these substances are somewhat higher than they are in temperate marine species. The balance of the depression, and ultimately the survival of the Antarctic fishes, rests on eight different antifreeze molecules, which are found in the body fluids of most notothenioid species studied to date. The full complement typically appears in most of the body fluids, except in the urine and the ocular fluid and within the cytoplasm of most cells, and it accounts for 3.5 percent of the weight of the fluids.

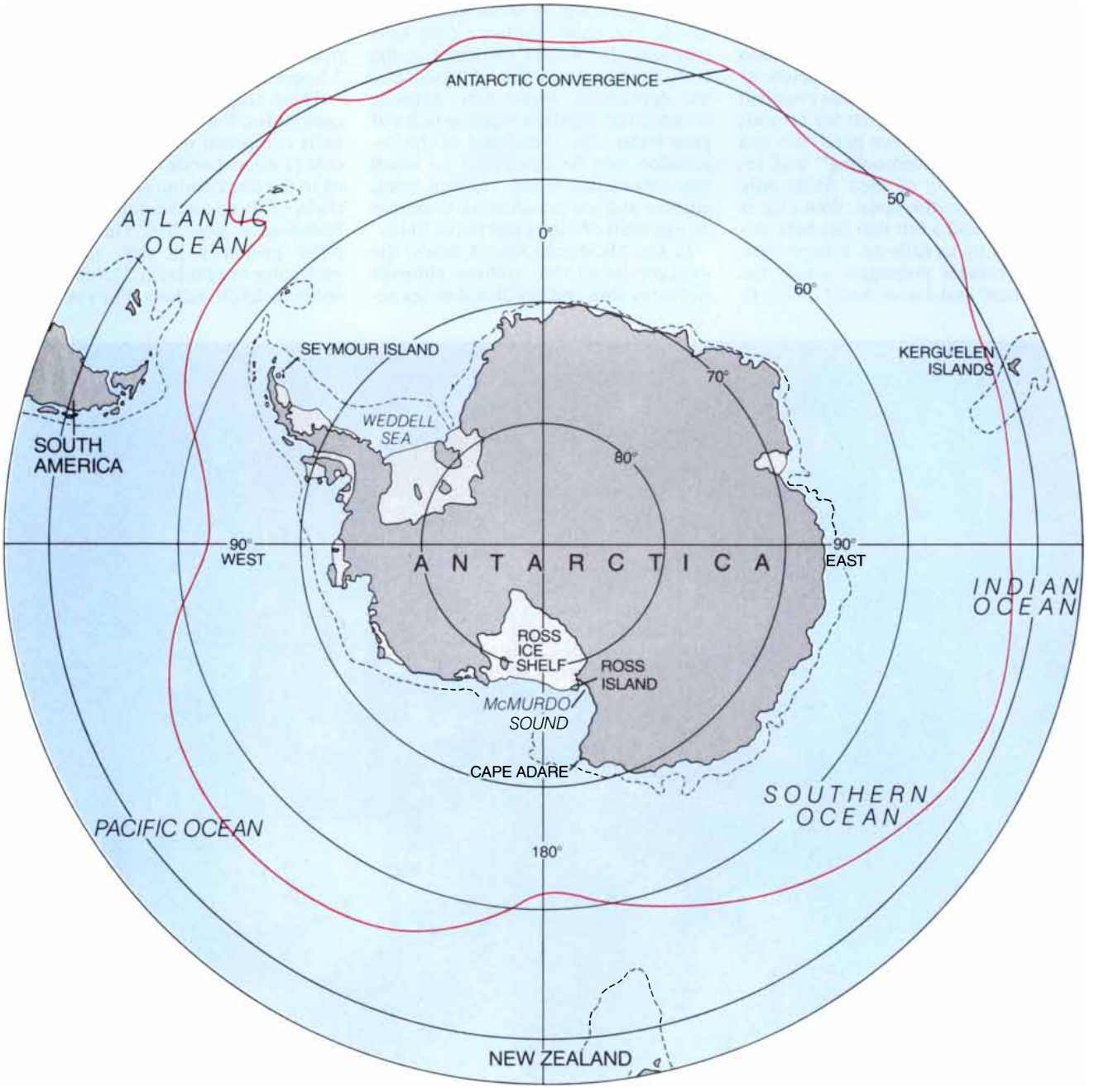
These antifreeze molecules are glycopeptides. Each consists of repeating units composed of a two-sugar molecule (a disaccharide) covalently bonded to the third amino acid of a peptide chain of three amino acids [*see top illustration on page 109*]. The molecules differ primarily in size, ranging in molecular weight from 33,700 daltons down to 2,600 daltons. For clarity ev-



TYPICAL NOTOTHENIOID, ***Trematomus nicolai,*** **lives, feeds and reproduces near the sea floor. The suborder Notothenioidei, a group of bony perchlike fishes, dominates in the Antarctic, accounting for some 90 percent of the individual fish in the region.**

ery glycopeptide has been assigned a number on the basis of its molecular weight; the largest molecule is No. 1 and the smallest molecule is No. 8. In glycopeptides 1 through 5 the amino acid sequence is alanine-alanine-threonine, and in the other glycopeptides the amino acid proline substitutes for some of the alanines. The antifreeze activity of the eight compounds increases with weight, and all the molecules appear to function similarly.

Notothenioid antifreezes lower the freezing point of fluids in a way that can best be understood by comparing their mode of action with that of more typical body-fluid solutes, such as glucose and sodium chloride. The freezing point of most solutions depends on their "colligative" properties, that is, on the number of solute particles present rather than on the nature of the particles. The more particles there are, the less likely it is that water molecules will aggregate and form an "embryonic" ice crystal. In water, sodium chloride depresses the freezing point nearly twice as much as glucose does because the salt dissociates into separate sodium and chloride ions. Glycopep-



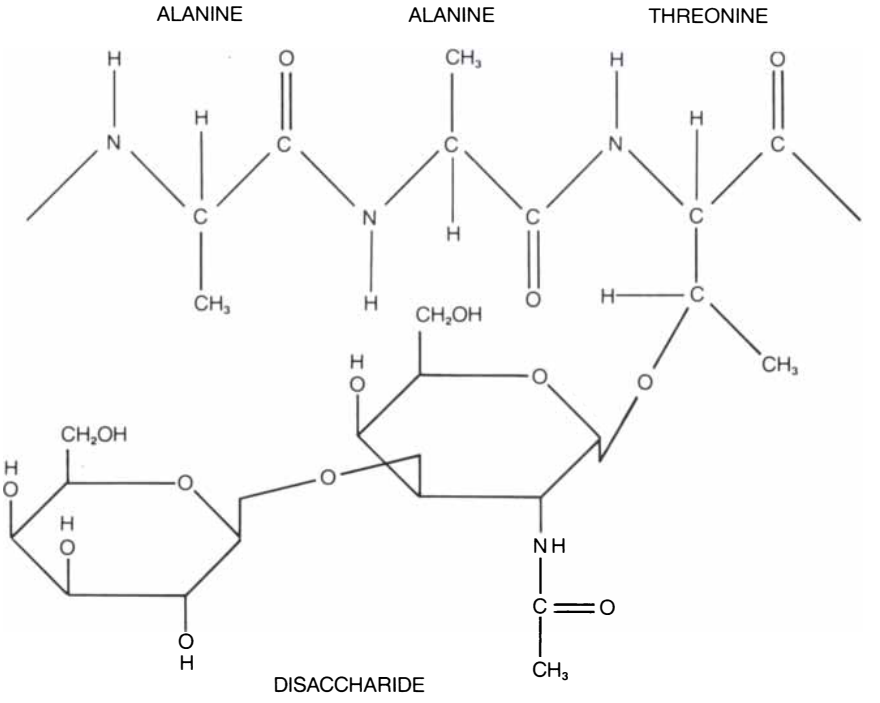
OCEANOGRAPHIC FEATURES that may have influenced the evolution of the notothenioid fishes include a narrow, deep continental shelf and the Antarctic Convergence, a zone of abrupt thermal change (*red line*). The water close to the margin of Antarctica reaches a depth of 1,000 meters (*broken line*), leaving little suitable habitat for many shallow-water fishes that might otherwise compete with the notothenioid fishes. The Antarctic Convergence, which forms the northern boundary of the "Southern Ocean" (where the Atlantic, Pacific and Indian oceans mingle), prevents warm surface water from flowing south into the Antarctic region. The development of the convergence may have contributed to the cooling of the Southern Ocean and hence to the evolution of the notothenioids under the influence of the cold. Today the temperature of the Southern Ocean rarely rises above two degrees Celsius.

tide antifreezes, in contrast, act noncolligatively: they can lower the freezing point of body fluids from 200 to 300 times more than would be expected on the basis of particle number. (The melting point also is lowered, but slightly and in a colligative manner.)

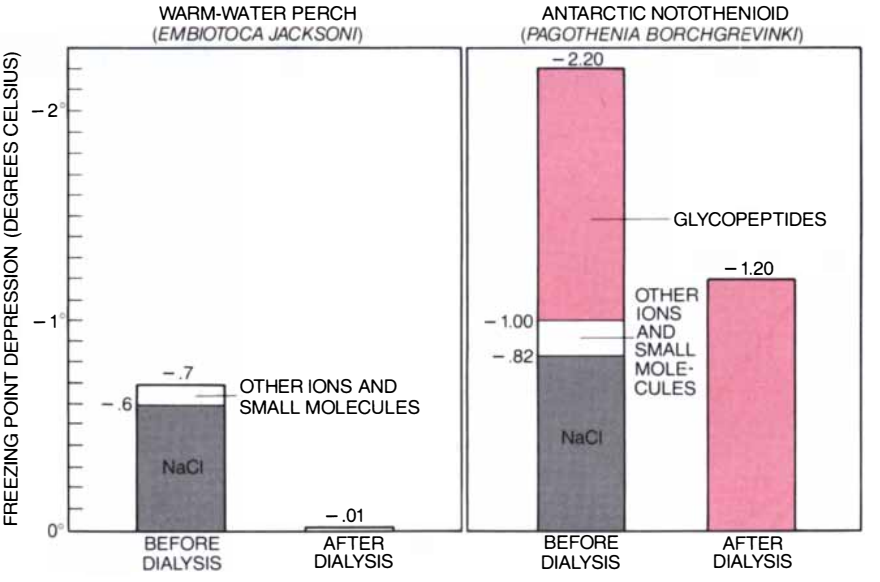
By what mechanism might these noncolligative antifreezes keep McMurdo Sound fishes from freezing in ice-laden waters? Chemists have long known that adsorbed impurities can inhibit the growth of small crystals and that, for unclear reasons, impurities consisting of a large number of repeating molecular subunits are particularly effective in this regard. One of us (DeVries) therefore suspected that the glycopeptide antifreezes might safeguard Antarctic notothenioids by adsorbing to minute ice crystals and so preventing their growth. Studies by DeVries and his students John G. Duman and James A. Raymond, then at the Scripps Institution of Oceanography, support this notion. The group found that the glycopeptides do indeed adsorb to ice while it is forming.

Although the events occurring at the molecular level are difficult to view, the group has proposed a likely description. We believe that ice cannot propagate over adsorbed glycopeptide molecules and is forced to grow only in the small spaces between them [*see illustration on page 112*]. Furthermore, these advancing fronts are curved and consequently have a large surface area in relation to their volume. Such ice fronts ultimately lose water molecules to the surrounding liquid, a phenomenon that halts their growth. In order for water molecules to be added to the fronts, the temperature of this surrounding liquid would have to be lowered. In other words, the development of highly curved fronts depresses the freezing point of the liquid.

Although this hypothesis could be interpreted as indicating that ice crystals form in the body fluids and are then stunted by a barrage of antifreeze molecules, recent experiments suggest otherwise. When notothenioids were placed in ice-free water, they did not freeze until their temperature fell to −6 degrees. Apparently they do not usually develop ice crystals inside the body unless ice first penetrates from the outside. The major threat to the notothenioids' ability to survive the cold must therefore be external ice, and the main role of the glycopeptides is probably to keep the ice from propagating across the integument. Some additional evidence is consistent with this suggestion. When the interior surface of scaleless skin is bathed with



BASIC STRUCTURAL SUBUNIT of notothenioid antifreeze compounds consists of a disaccharide (two-sugar) molecule linked to the third amino acid of a tripeptide (three-amino-acid) chain. The antifreeze molecules, known as glycopeptides, consist of repeating subunits and are identified by number, according to their molecular weight. (No. 1 is the heaviest molecule, at 33,700 daltons; No. 8 is the lightest, at 2,600 daltons.) In glycopeptides 1 through 5 the subunit peptide chain consists of alanine-alanine-threonine, as is shown; in glycopeptides 6 through 8 proline replaces some of the alanines (not shown).



COMPARISONS of blood plasma drawn from a Pacific perch (a warm-water fish) and from an Antarctic notothenioid highlight differences in the freezing points and composition of the plasma. Workers compared the freezing points of the plasma before and after it was dialyzed, or filtered, to remove solutes smaller than 1,000 daltons. Before dialysis the freezing point of the perch plasma was −.7 degree Celsius; after dialysis it rose to −.01 degree, essentially the freezing point of pure water (0 degrees). This change indicates that sodium chloride and other small solutes normally account for the mild freezing point depression (below zero) of perch plasma. Before dialysis the freezing point of notothenioid blood plasma was −2.20 degrees; after dialysis it rose to −1.20 degrees. The presence of a continued freezing-point depression after dialysis indicates that the glycopeptide molecules in notothenioid blood, which are larger than 1,000 daltons, account for a significant part of the typical freezing-point depression of that blood; in fact, they account for about half of the freezing-point depression that has been found.

a salt solution containing antifreeze molecules, the skin acts as a barrier to ice propagation from the outside. When the antifreeze molecules are not present, ice propagates readily across the integument.

We do not yet know how antifreeze molecules bind to ice, in part because no one fully understands their three-dimensional structure in solution. We do know that the hydroxyl (–OH) groups and other polar groups of the antifreezes branch out from the backbone. We also think these groups are binding sites. Indeed, we have shown that the hydroxyl groups in the sugar fraction of the glycopeptides are apparently vital to the antifreeze function of the molecules. When these hydroxyls are experimentally inactivated (by adding an acetyl group, $CH_3CO–$), the glycopeptide molecules lose their antifreeze effect.

The polar groups can potentially form hydrogen bonds with water molecules in the ice lattice, which are arranged in rows of hexagons that have oxygen atoms at the "corners." For maximal hydrogen bonding to occur between the ice lattice and the polar groups in the glycopeptide, a lattice match would have to be possible; that is, the groups would have to be separated by distances that correspond to the distances between oxygen atoms on a growing front of the ice lattice.

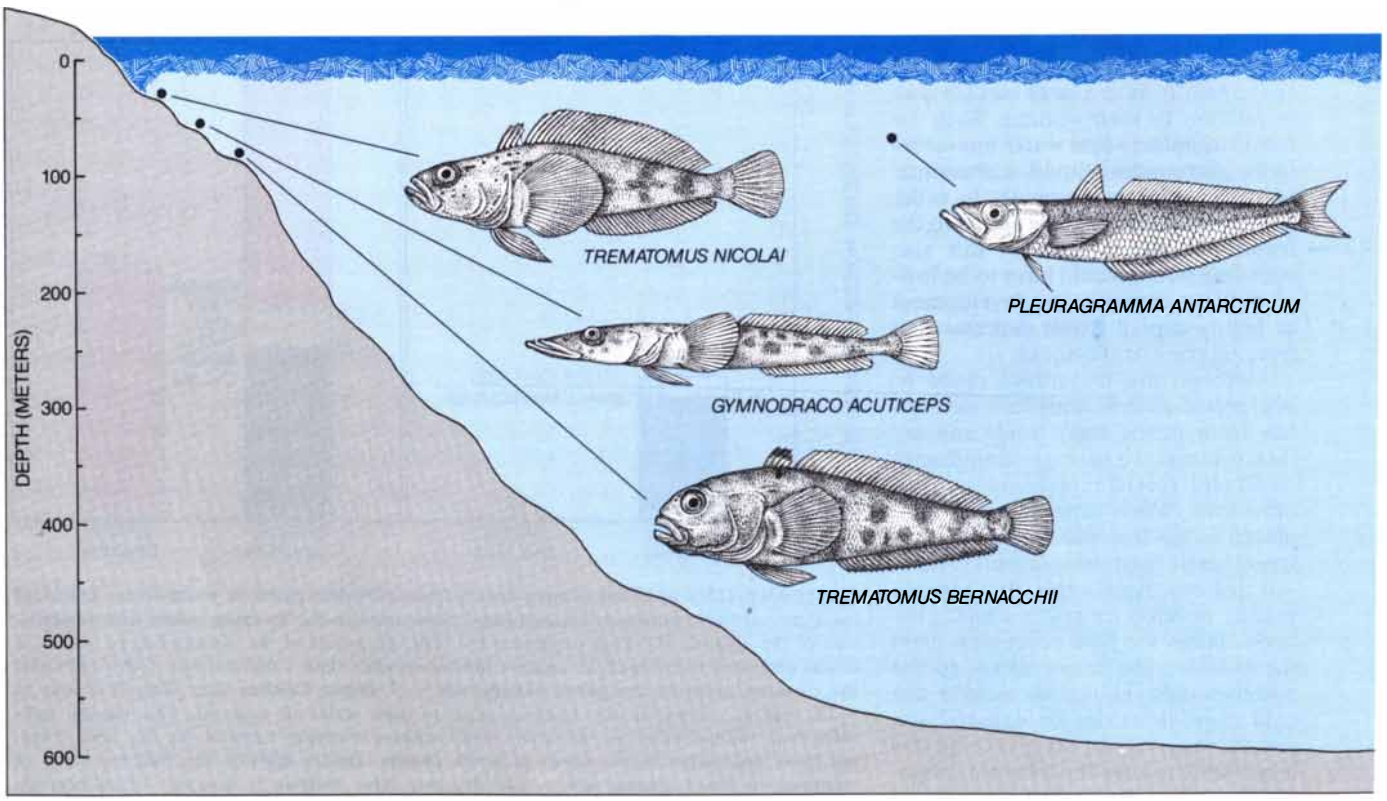
This ideal requirement may be met in reality. Models of glycopeptides reveal that many of the hydroxyls of the sugar side chains are spaced 4.5 angstrom units apart—just about the distance separating some of the oxygen atoms on the horizontal plane of the ice lattice.

Another way antifreezes might bind to ice is through the carbonyl (–CO–) groups in the amino acid chain. It could be that the sugar side chains of the glycopeptides somehow keep the polypeptide backbone of the molecule in an expanded conformation. Under this condition every other carbonyl group would project from the same side of the polypeptide and be separated by about 7.3 angstroms. A similar distance separates certain oxygens in the ice lattice.

Regardless of their mechanism of action, the antifreezes of the McMurdo Sound notothenioids are clearly needed all year round. This raises an interesting problem. Antarctic fishes have limited energy stores and must conserve energy, particularly in the austral winter, when the ecosystem's productivity is particularly low. How then do they maintain an adequate supply of antifreezes without expending great energy on synthesis?

The fact that notothenioids have no antifreeze molecules in their urine suggests that some mechanism associated with the kidneys may play a role in conserving antifreezes. In most vertebrates the glycopeptide antifreezes, which are relatively small molecules, would be expected to escape routinely into the urine through the kidney's glomeruli: the tufts of capillaries that act as blood filters. The pressure inside the glomeruli usually forces molecules smaller than 40,000 daltons to pass from the blood into the system of tubules that collect the urine. A fish with glomeruli could theoretically retrieve the antifreezes before the compounds leave the tubules, by breaking the molecules down into their smallest components (amino acids and sugars), returning the components to the blood from the urine and then resynthesizing the glycopeptides. The cost, however, would be high: two energy units for each bond that is formed between two amino acids.

In 1972 DeVries and his student Gary H. Dobbs III, then at the Scripps Institution, examined kidney tissues



SEVEN SPECIES of notothenioids are among some 14 fish species in McMurdo Sound and some 90 in the Southern Ocean. Many notothenioids, such as *Trematomus nicolai, T. bernacchii, T. loennbergii* and *Gymnodraco acuticeps,* dwell on or near the sea floor, albeit at different depths. (The dots indicate typical levels; some fishes have a wider range than is shown.) A few species have

of notothenioids under a microscope. They found that 12 out of 12 species lack glomeruli. Moreover, studies of antifreeze molecules labeled with radioactive isotopes have demonstrated that the kidneys prevent the glycopeptides from ever entering the urine. These aglomerular species produce urine by a secretory process. Cells that line the walls of tubules draw only selected wastes from the blood, leaving the needed antifreezes in circulation. This eliminates the need for resynthesis and thereby saves energy.

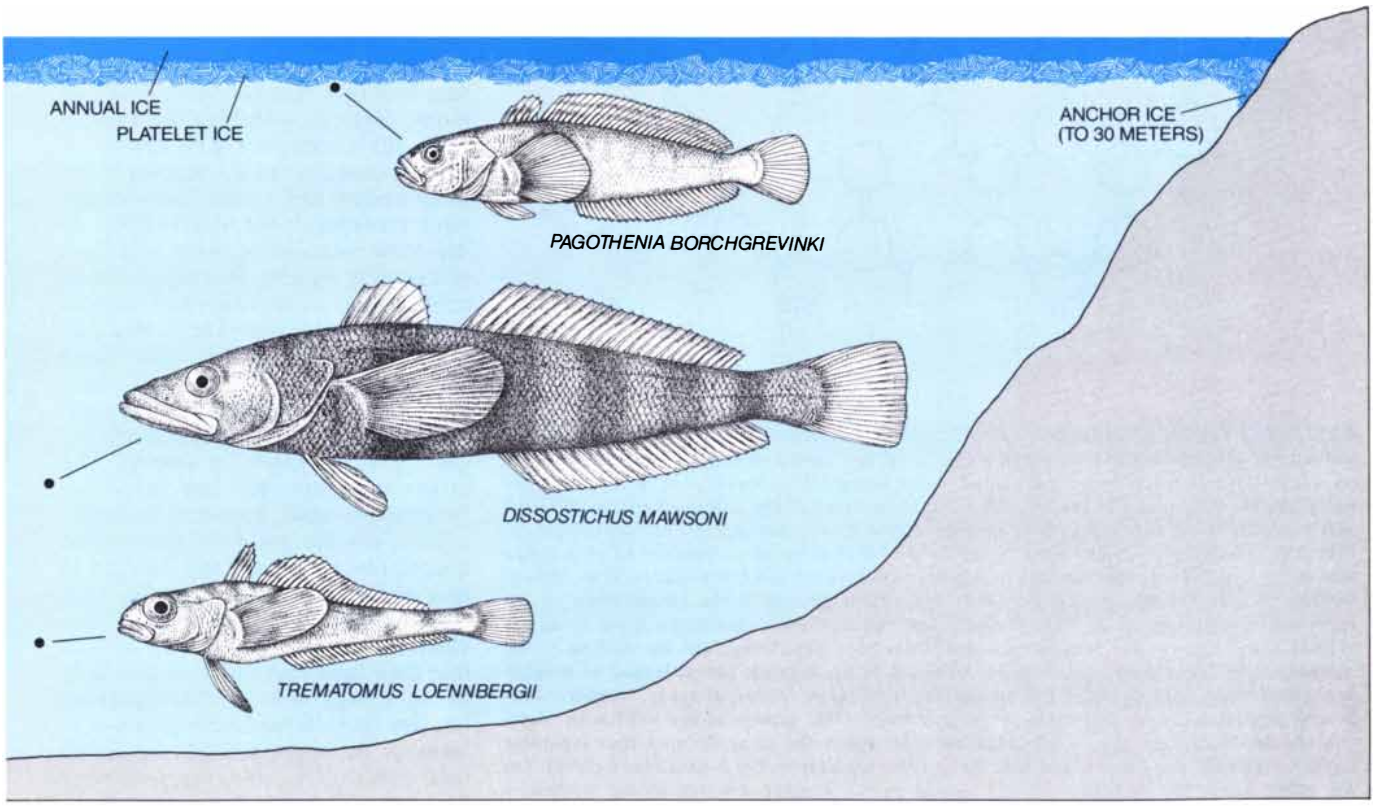
We are not certain that the aglomerular condition of the kidneys in McMurdo Sound notothenioids evolved specifically to conserve antifreezes and energy, but the kidney's importance to those functions is consistent with such a hypothesis. Moreover, we have recently discovered that the New Zealand thornfish (*Bovichthys variegatus*), one of a few notothenioid species found in temperate waters, has many glomeruli. Because the thornfish is among the least specialized of all notothenioids, the fact that it has glomeruli suggests the aglomerular condition of other notothenioids is a specialized adaptation, one that could conceivably have developed specifically as an aid to cold-water survival.

The presence of neutral buoyancy in at least two notothenioid species is another example of an evolutionary adaptation that has enabled certain notothenioids to conserve energy. We discovered this feature about 10 years ago when we captured several specimens of the largest known notothenioid: the Antarctic toothfish (*Dissostichus mawsoni*). (Whereas notothenioids are generally from 15 to 30 centimeters long, the typical toothfish is 127 centimeters long and weighs 28 kilograms.) We had set our fishing line between 300 and 500 meters deep, and we initially assumed the fish rose from their habitat near the sea floor to strike our hooks. We developed some doubts, however, when we noted that *Dissostichus,* as well as *Pleuragramma antarcticum,* the smaller notothenioid species that accounted for 70 percent of the stomach contents of *Dissostichus,* had the streamlined appearance of mid-water fishes. Perhaps *Dissostichus* and its prey were highly modified, mid-water offshoots of the bottom-dwelling notothenioid stock.

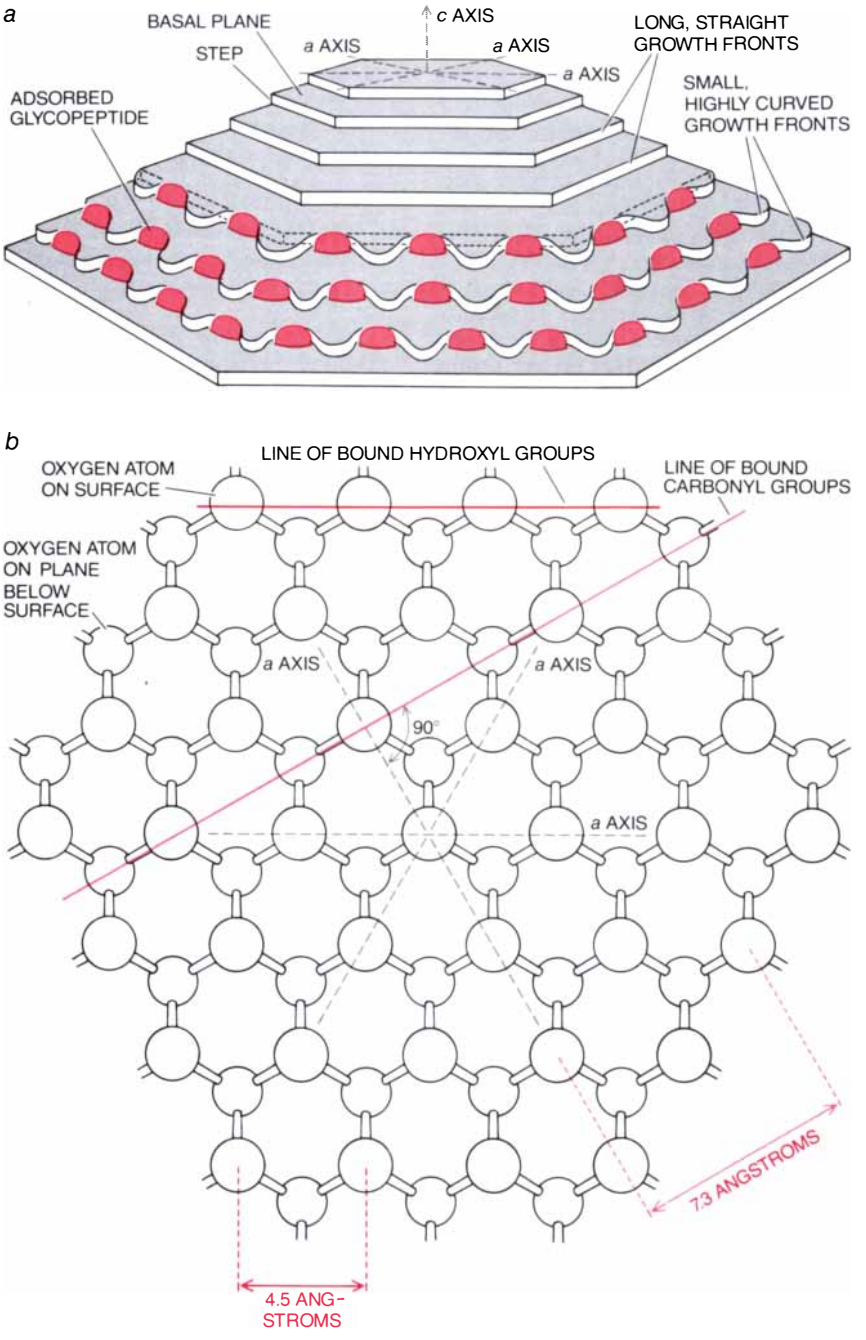
In 1978 we began a series of investigations to determine whether these two species might be buoyant enough to live efficiently in middle-level waters and, if they were, which of their adaptations made this buoyancy possible. Neutral buoyancy is actually a comparative measure, derived by dividing a specimen's weight in the water by its weight outside the water and multiplying by 100. The closer the result is to zero, the closer the fish is to being neutrally buoyant. The mean result for *Dissostichus* was .01 and for *Pleuragramma* .6, close enough to zero for these species to be considered neutrally buoyant.

Simple dissection confirmed our belief that these fishes, like their bottom-dwelling relatives, lack a swim bladder, the gas-filled sac that typically results in neutral buoyancy in fish. Something else had to account for our findings. One obvious possibility was a reduction in the bony material of the skeleton, typically the body's densest component. We pursued this idea with the help of an ordinary chef's knife from the mess hall. The knife easily cut through the skull of *Dissostichus,* which consisted largely of cartilage, as did the tail skeleton and the pectoral girdle, another major bone. Cartilage is considerably less dense than bone, offering a major saving in weight.

To quantify the degree of mineralization of the skeleton, we next employed a furnace to "ash" the skel-



radiated from the bottom, notably *Dissostichus mawsoni* and *Pleuragramma antarcticum,* which are mid-water fishes that live at depths ranging from the ones shown to 500 meters, and *Pagothenia borchgrevinki,* which is adapted to life in and under the platelet ice. *Dissostichus,* shown as larger than the other fishes, is six times as long and 250 times as heavy as most notothenioids.



GLYCOPEPTIDES ADSORBED TO ICE probably impede crystal growth by inhibiting the normal propagation of ice-growth steps, or layers, across the crystal's surface (*a*). An ice crystal grows when water molecules in the surrounding liquid join at steps on the basal (horizontal) plane of the crystal. Such steps start at the center of the basal plane and normally grow outward in long straight-edged fronts (*top layers*). In contrast, steps that meet glycopeptide antifreezes bound to the basal plane of an underlying step divide into many small fronts that become highly curved. Such curved fronts have a high ratio of surface area to volume, a condition that halts their growth if the temperature of the surrounding liquid does not fall. Exactly how the antifreeze molecules bind to an ice crystal is not known, but two likely possibilities have been suggested for binding to the surface of the basal plane (*b*). Water molecules in ice crystals are arranged in roughly hexagonal rings, each of which has an oxygen atom (*open circles*) at every "corner." Molecular models indicate that some of the hydroxl (–OH) groups of the antifreeze disaccharide side chains are spaced 4.5 angstrom units apart, the same distance that separates oxygen atoms on the basal plane that lie in rows parallel to the *a* axes (*dark color*). On the other hand, the carbonyl (–CO–) groups in the peptide fraction of the antifreeze molecules might also bind to the crystal (*light color*). According to molecular models, in glycopeptide antifreezes that have a completely extended conformation the distance separating every other carbonyl group is roughly 7.3 angstroms, about the distance between oxygen atoms on the basal plane that lie in rows running perpendicular to the *a* axes.

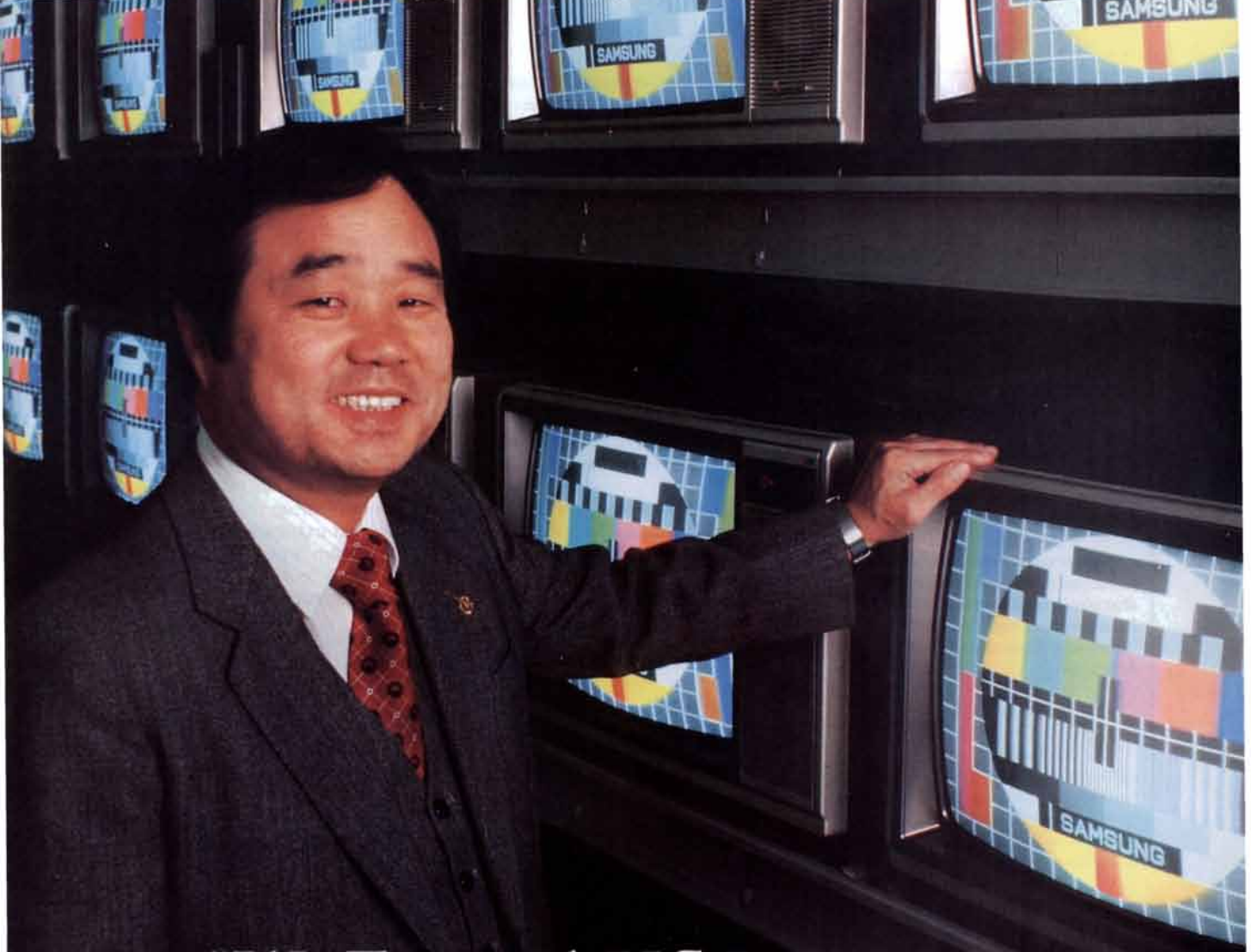
etons of *Dissostichus, Pleuragramma* and a bottom dweller, *Bovichthys.* This procedure vaporizes organic matter, leaving only the mineral residue of the skeleton. The ashed skeleton constituted .6 percent of the body weight of *Dissostichus* and only .3 percent of the body weight of *Pleuragramma,* in both cases a marked contrast to the 3.8 percent of body weight attributed to the ashed skeleton of *Bovichthys.*

We were not surprised by these results, but we were by our next discovery. An air bubble we had somehow introduced into the vertebral column of a partially dissected *Pleuragramma* specimen moved up and down the interior of the vertebrae when we tilted the fish like a seesaw. In most fishes such movement is not possible: the vertebral column is solid bone and represents the largest part of the skeleton by weight. Further inspection of *Pleuragramma* revealed that its vertebrae are essentially hollow: a thin collar of bone surrounds a gelatinous embryonic structure (the notochord) that persists into adult life.

In addition to a reduction in bone, an abundance of triglyceride (a type of lipid, or fat, that is less dense than McMurdo Sound seawater) would also contribute to buoyancy. Both *Dissostichus* and *Pleuragramma* have triglyceride deposits, although the deposits take different forms.

A cross section of *Dissostichus* glistens with lipid that fills fat cells in two major deposits: a blubberlike, two- to eight-millimeter thick layer under the skin (accounting for 4.7 percent of the body weight) and a more dispersed deposit throughout the muscle fibers of the trunk (accounting for 4.8 percent of the body weight). *Pleuragramma,* in contrast, has a unique method of lipid retention for a vertebrate: it stores its lipid in sacs rather than in cells. These sacs are from .2 millimeter to three millimeters in diameter and are abundant under the skin in the pectoral region, near the center of gravity. The largest sacs, however, are found between the muscles, deeper in the body.

We have not yet fully determined the relative benefits of storing lipid in sacs as opposed to in fat cells. We originally thought that lipid sacs were exclusively designed for buoyancy and that their lipid stores were unavailable as an energy source; *Pleuragramma* has few fat cells, and such cells would seem to be important for regulating lipid removal. More recently we have found that muscle cells adjacent to the sacs have large vacuoles, or membrane-bound cavities, that may contain triglyceride. Perhaps the muscles



"WE FOUND A U.S. LOCATION THAT WASN'T FOREIGN TO OUR BUSINESS."

– Hai-Min Lee, *President* Samsung International, Inc.

Samsung Electronics, a subsidiary of a South Korean multinational corporation has been doing business in America for years. When they needed to locate their new $25 million TV and Microwave manufacturing plant in America, the decision was clear.

They chose New Jersey.

"Because I chose a State that wasn't foreign to my business, I was in business by the end of the year."

"The Morris County site was chosen after looking at locations in several other states," says Hai-Min Lee, "The willingness of New Jersey to develop an attractive $9.1 million bond and financing program made the decision for us. We've also added to our N.J. commitment with a $10 million, 8 year lease on office and warehouse facilities in Bergen County."

Governor Thomas H. Kean said, "Samsung discovered our flexible commitment to business that has attracted so many other successful companies to New Jersey." For more information write Borden R. Putnam, Commissioner, Department of Commerce & Economic Development Box CN820, Trenton NJ, 08625. Or phone (609) 292-2444.

"What more could we ask for," says Hai-Min Lee, "As soon as we spoke to New Jersey we felt right at home."

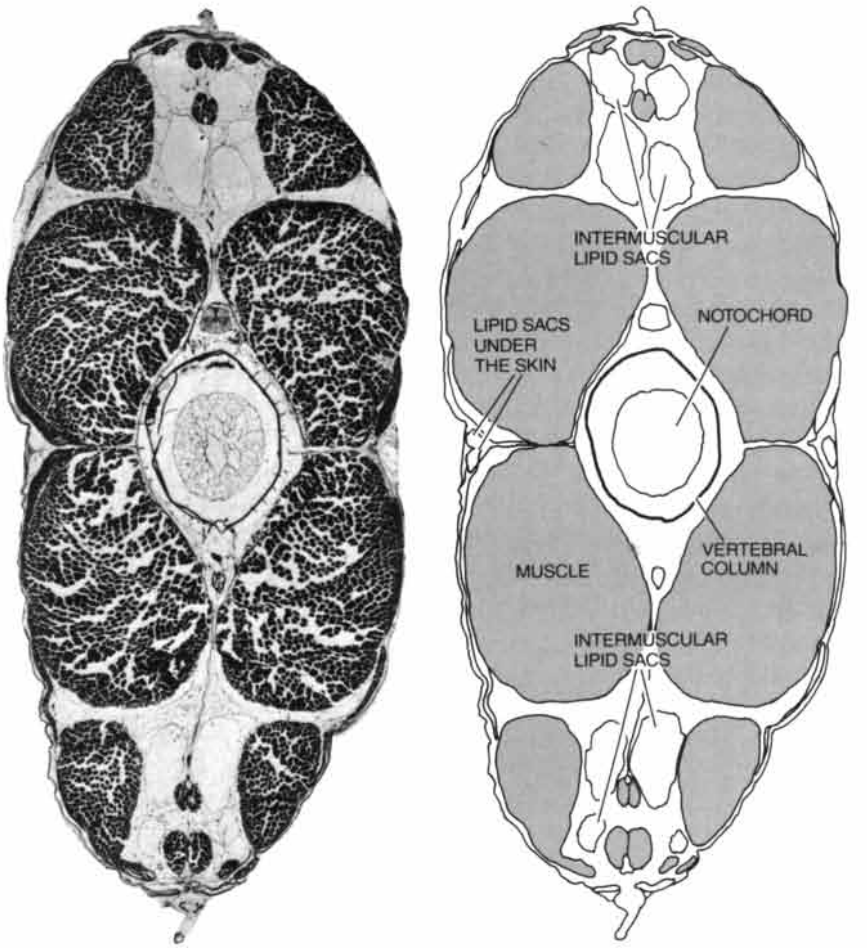
NJ & You
Perfect Together

NEW JERSEY DEPARTMENT OF COMMERCE & ECONOMIC DEVELOPMENT
Thomas H. Kean, *Governor*, Borden R. Putnam, *Commissioner*

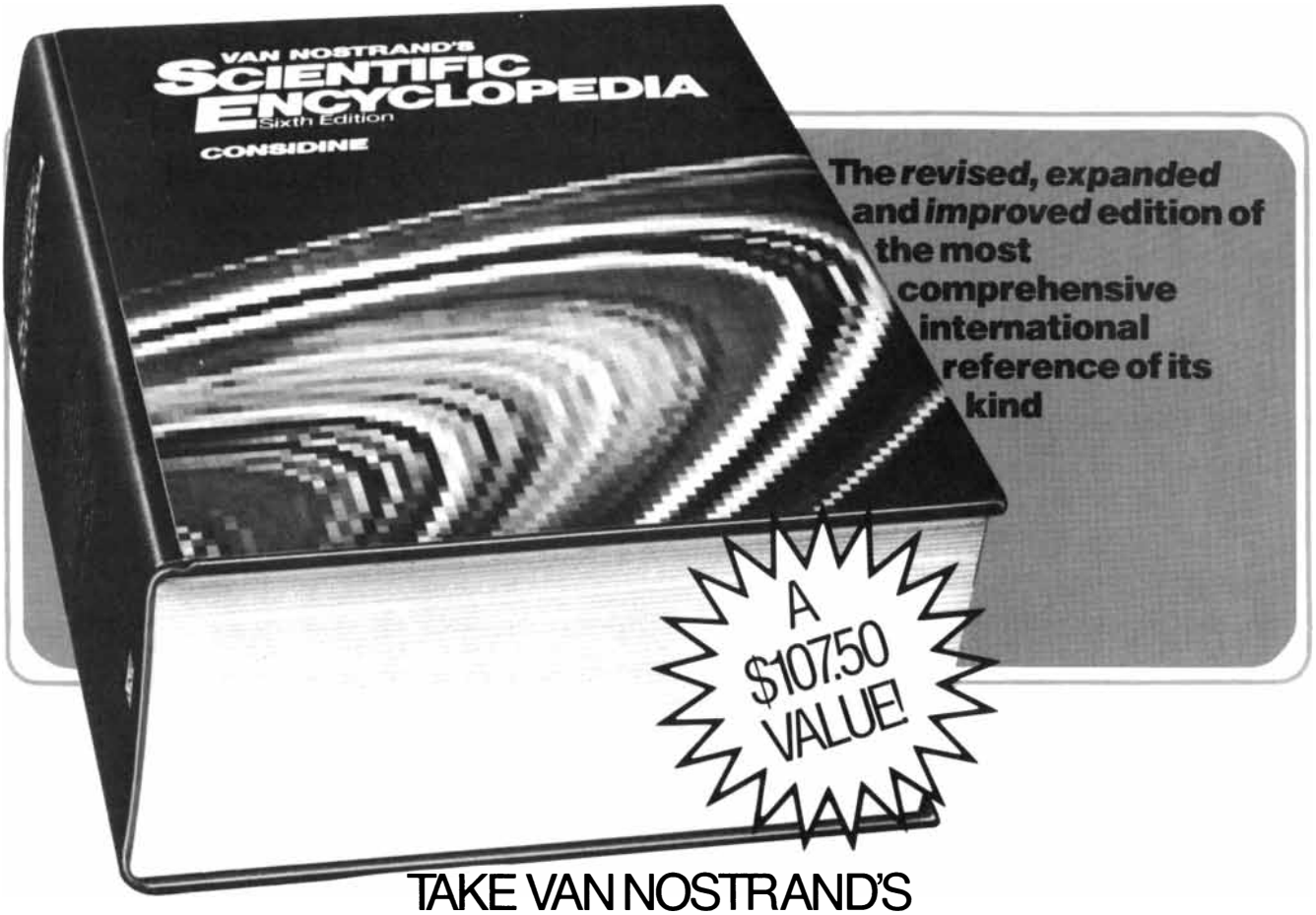
add lipid to the sacs for buoyancy and also remove them when they are needed for energy. We intend to test this hypothesis on our next trip to the McMurdo Sound research station.

Skeletal reduction and lipid deposition together can account for the reduced density and the neutral buoyancy of both *Dissostichus* and *Pleuragramma*. In turn weightlessness certainly facilitates their exploitation of the underutilized mid-waters of the Antarctic. Indeed, both species are abundant and widely distributed there. *Pleuragramma,* for instance, is known to be the dominant species in the Ross and Weddell seas, as well as in the waters of McMurdo Sound.

The adaptations that have enabled notothenioids to radiate into the mid-waters fascinate us from an anatomical point of view; we are also struck by their ecological implications. It is time to revise the standard accounts of the mid-water food web. As it is usually described, the web is short, including only planktonic plant life, krill (a shrimplike crustacean), seals and whales. We and other investigators have shown, however, that the prevailing vision of the composition and complexity of the food web should be expanded to include neutrally buoyant notothenioids and perhaps other organisms as well. Indeed, in areas of the Antarctic Ocean where krill is sparse *Pleuragramma* may take the place of krill in the food chain. Finally, from yet another point of view, we hope our studies of buoyancy and antifreezes provide a glimpse of the extraordinarily wide scope of adaptation and evolution that has occurred at the coldest extreme of the environmental spectrum, a place that was once widely thought to be incompatible with significant marine life.



CROSS SECTION of a *Pleuragramma* reveals features that contribute to neutral buoyancy, or weightlessness in water. Neutral buoyancy enables *Pleuragramma* to live in the mid-waters without devoting energy to flotation. One prominent feature is the presence of lipid sacs. Lipid, or fat, is not as dense as seawater and so promotes flotation. The notochord within *Pleuragramma*'s vertebral column also contributes to neutral buoyancy because it is lightweight; in most bony fishes the vertebrae are solid bone and very heavy. (Fixation for microscopy has caused the notochord to shrink quite far from the bone.)



TAKE VAN NOSTRAND'S

SCIENTIFIC ENCYCLOPEDIA

for only $19.95!

When you join **The Library of Science**—*the oldest and most respected science book club in the world.* You simply agree to buy 3 more books—at handsome discounts—in the next 12 months.

For over 40 years, this unique single-volume sourcebook has been acclaimed for the breadth and depth of its knowledge, accuracy, clarity and practicality. Now its coverage has been dramatically expanded to bring you comprehensive information on *7,300 entries,* including the latest discoveries in science and technology.

And to give you "instant access" to this vital data, the format of the SCIENTIFIC ENCYCLOPEDIA has been revised so it's easier than ever before to read and cross reference. You'll learn terms and definitions; theories and their practical applications—on a vast range of topics such as:

- animal life
- astronomy
- biosciences
- chemistry
- computer science
- earth science
- energy sources
- mathematics
- medicine and anatomy
- physics
- plant sciences

This entire 3,100 page ENCYCLOPEDIA—a $107.50 value—can be yours for only $19.95.

What's *new* about the new edition of the SCIENTIFIC ENCYCLOPEDIA?

- 500 additional pages, for a total count of 3,100
- Comprehensive coverage of 7,300 topics
- 2,550 photographs, diagrams and charts, plus 600 tables to give you vital details at a glance
- Contributions by over 200 international experts
- Over 2.3 million words
- Extra-sturdy binding for durability

About The Library of Science...

The Library of Science is the nation's oldest and most respected science book club. Dedicated to serving both professionals and serious amateurs, it surveys the entire field of published sources on scientific subjects to make the very finest works available to club members at substantial savings.

MEMBERSHIP BENEFITS • In addition to getting VAN NOSTRAND'S SCIENTIFIC ENCYCLOPEDIA for only $19.95 when you join, you keep saving substantially on the books you buy—up to 30% and occasionally even more. • Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings up to 65% off publishers' prices. • At 3–4 week intervals (16 times per year), you will receive the Library of Science News, describing the coming Main Selection and Alternate Selections together with a dated reply card. • If you want the Main Selection, do nothing and it will be sent to you automatically. • If you prefer another selection, or no book at all, simply indicate your choice on the card, and return it by the date specified. • You will have at least 10 days to decide. If, because of late mail delivery of the News, you should receive a book you do not want, we guarantee return postage.

Reply card removed? Please write: The Library of Science, Dept. 2-BL2-85755, Riverside, N.J. 08075 for membership information and an application.

Scientific American 11/86

Features and Objects in Visual Processing

The seemingly effortless ability to perceive meaningful wholes in the visual world depends on complex processes. The features automatically extracted from a scene are assembled into objects

by Anne Treisman

If you were magically deposited in an unknown city, your first impression would be of recognizable objects organized coherently in a meaningful framework. You would see buildings, people, cars and trees. You would not be aware of detecting colors, edges, movements and distances, and of assembling them into multidimensional wholes for which you could retrieve identities and labels from memory. In short, meaningful wholes seem to precede parts and properties, as the Gestalt psychologists emphasized many years ago.

This apparently effortless achievement, which you repeat innumerable times throughout your waking hours, is proving very difficult to understand or to simulate on a computer—much more difficult, in fact, than the understanding and simulation of tasks that most people find quite challenging, such as playing chess or solving problems in logic. The perception of meaningful wholes in the visual world apparently depends on complex operations to which a person has no conscious access, operations that can only be inferred on the basis of indirect evidence.

Nevertheless, some simple generalizations about visual information processing are beginning to emerge. One of them is a distinction between two levels of processing. Certain aspects of visual processing seem to be accomplished simultaneously (that is, for the entire visual field at once) and automatically (that is, without attention being focused on any one part of the visual field). Other aspects of visual processing seem to depend on focused attention and are done serially, or one at a time, as if a mental spotlight were being moved from one location to another.

In 1967 Ulric Neisser, then at the University of Pennsylvania, suggested that a "preattentive" level of visual processing segregates regions of a scene into figures and ground so that a subsequent, attentive level can identify particular objects. More recently David C. Marr, investigating computer simulation of vision at the Massachusetts Institute of Technology, found it necessary to establish a "primal sketch": a first stage of processing, in which the pattern of light reaching an array of receptors is converted into a coded description of lines, spots or edges and their locations, orientations and colors. The representation of surfaces and volumes and finally the identification of objects could begin only after this initial coding.

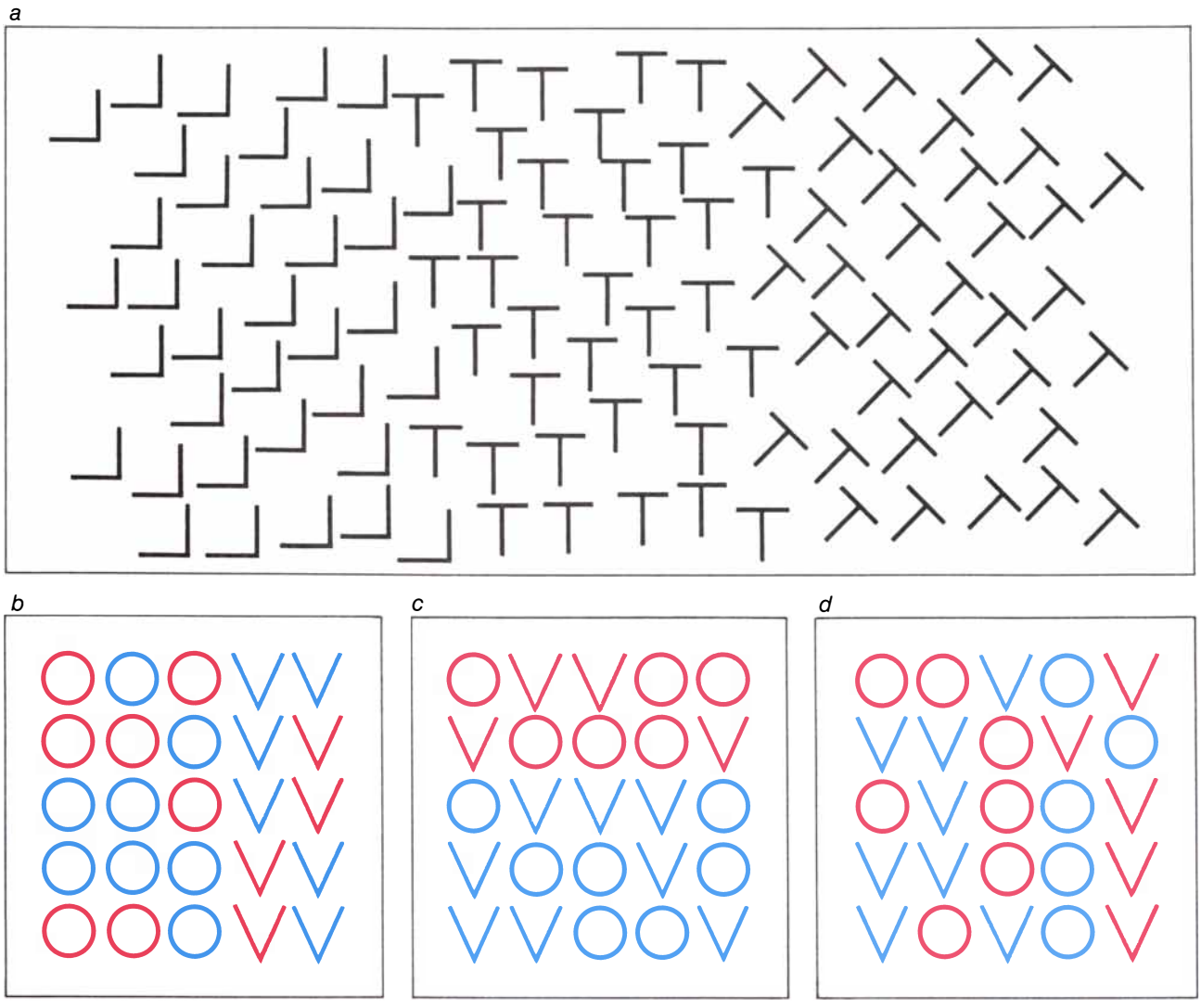
In brief, a model with two or more stages is gaining acceptance among psychologists, physiologists and computer scientists working in artificial intelligence. Its first stage might be described as the extraction of features from patterns of light; later stages are concerned with the identification of objects and their settings. The phrase "features and objects" is therefore a three-word characterization of the emerging hypothesis about the early stages of vision.

I think there are many reasons to agree that vision indeed applies specialized analyzers to decompose stimuli into parts and properties, and that extra operations are needed to specify their recombination into the correct wholes. In part the evidence is physiological and anatomical. In particular, the effort to trace what happens to sensory data suggests that the data are processed in different areas of considerable specialization. One area concerns itself mainly with the orientation of lines and edges, another with color, still another with directions of movement. Only after processing in these areas do data reach areas that appear to discriminate between complex natural objects.

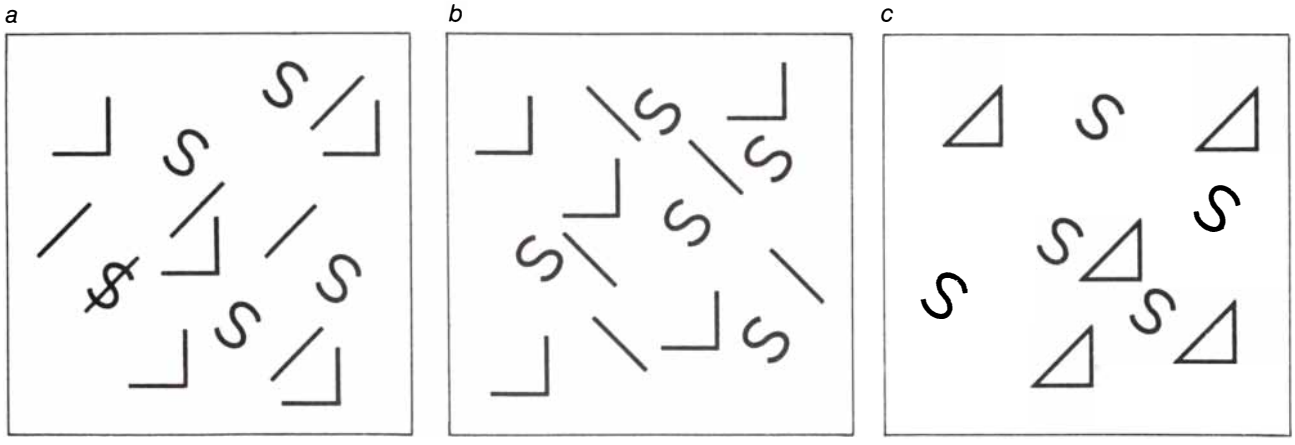
Some further evidence is behavioral. For example, it seems that visual adaptation (the visual system's tendency to become unresponsive to a sustained stimulus) occurs separately for different properties of a scene. If you stare at a waterfall for a few minutes and then look at the bank of the river, the bank will appear to flow in the opposite direction. It is as if the visual detectors had selectively adapted to a particular direction of motion independent of *what* is moving. The bank looks very different from the water, but it nonetheless shows the aftereffects of the adaptation process.

How can the preattentive aspect of visual processing be further subjected to laboratory examination? One strategy is suggested by the obvious fact that in the real world parts that belong to the same object tend to share prop-

PRIOR KNOWLEDGE AS A GUIDE in visual perception is tested by asking subjects to search for a familiar object in a photograph of an unexceptional scene (*top*) or in a jumbled photograph of the scene (*bottom*). Here the task is simply to find the bicycle. It tends to take longer in the jumbled image. The implication is that knowledge of the world (in this case expectations about the characteristic locations of bicycles in an urban landscape) speed up perception and make it less subject to error. Certain early aspects of the information processing that underlies visual perception nonetheless seem to happen automatically: without the influence of prior knowledge. The illustration was modeled after experiments done by Irving Biederman of the State University of New York at Buffalo.



BOUNDARIES THAT "POP OUT" of a scene are likely to reveal the simple properties, or features, of the visual world that are seized on by the initial stage of visual processing. For example, a boundary between T's and tilted T's pops out, whereas a boundary between T's and L's does not (*a*). The implication is that line orientations are important features in early visual processing but that particular arrangements of conjunctions of lines are not. A boundary between O's and V's pops out (*b*). The implication is that simple shape properties (such as line curvature) are important. A boundary between red and blue shapes pops out (*c*), implying that color is important. A boundary between conjunctions of shape and color, in this case red V's and blue O's versus red O's and blue V's (*d*), does not pop out. Evidently early vision deals only with individual features, not with conjunctions of features.



ILLUSORY DOLLAR SIGNS are an instance of false conjunctions of features. Subjects were asked to look for dollar signs in the midst of S's and line segments (*a*). They often reported seeing the signs when the displays to which they were briefly exposed contained none (*b*). They had the same experience about as often when the line segment needed to complete a sign was embedded in a triangle (*c*). The experiment suggests that early visual processing can detect the presence of features independent of location.

erties: they have the same color and texture, their boundaries show a continuity of lines or curves, they move together, they are at roughly the same distance from the eye. Accordingly the investigator can ask subjects to locate the boundaries between regions in various visual displays and thus can learn what properties make a boundary immediately salient—make it "pop out" of a scene. These properties are likely to be the ones the visual system normally employs in its initial task of segregating figure from ground.

It turns out that boundaries are salient between elements that differ in simple properties such as color, brightness and line orientation but not between elements that differ in how their properties are combined or arranged [*see top illustration on opposite page*]. For example, a region of T's segregates well from a region of tilted T's but not from a region of L's made of the same components as the T's (a horizontal line and a vertical line). By the same token, a mixture of blue V's and red O's does not segregate from a mixture of red V's and blue O's. It seems that the early "parsing" of the visual field is mediated by separate properties, not by particular combinations of properties. That is, analysis of properties and parts precedes their synthesis. And if parts or properties are identified before they are conjoined with objects, they must have some independent psychological existence.

This leads to a strong prediction, which is that errors of synthesis should sometimes take place. In other words, subjects should sometimes see illusory conjunctions of parts or properties drawn from different areas of the visual field. In certain conditions such illusions take place frequently. In one experiment my colleagues and I flashed three colored letters, say a blue X, a green T and a red O, for a brief period (200 milliseconds, or a fifth of a second) and diverted our subjects' attention by asking them to report first a digit shown at each side of the display and only then the colored letters. In about one trial in three the subjects reported the wrong combinations—perhaps a red X, a green O or a blue T.

The subjects made these conjunction errors much more often than they reported a color or shape that was not present in the display, which suggests that the errors reflect genuine exchanges of properties rather than simply misperceptions of a single object. Many of these errors appear to be real illusions, so convincing that subjects demand to see the display again to convince themselves that the errors were indeed mistakes.

We have looked for constraints on the occurrence of such illusory conjunctions. For example, we have asked whether objects must be similar for their properties to be exchanged. It seems they do not: subjects exchanged colors between a small, red outline of a triangle and a large, solid blue circle just as readily as they exchanged colors between two small outline triangles. It is as if the red color of the triangle were represented by an abstract code for red rather than being incorporated into a kind of analogue of the triangle that also encodes the object's size and shape.

We also asked if it would be harder to create illusory conjunctions by detaching a part from a simple unitary shape, such as a triangle, than by moving a loose line. The answer again was no. Our subjects saw illusory dollar signs in a display of S's and lines. They also saw the illusory signs in a display of S's and triangles in which each triangle incorporated the line the illusion required [*see bottom illustration on opposite page*]. In conscious experience the triangle looks like a cohesive whole. Nevertheless, at the preattentive level its component lines seem to be detected independently.

To be sure, the triangle may have an additional feature, namely the fact that its constituent lines enclose an area, and this property of closure might be detected preattentively. If so, the perception of a triangle might require the detection of its three component lines in the correct orientations and also the detection of closure. We should then find that subjects do not see illusory triangles when they are given only the triangles' separate lines in the proper orientations. They may need a further stimulus, a different closed shape (perhaps a circle), in order to assemble illusory triangles. That is indeed what we found.

Another way to make the early, preattentive level of visual processing the subject of laboratory investigation is to assign visual-search tasks. That is, we ask subjects to find a target item in the midst of other, "distractor" items. The assumption is that if the preattentive processing occurs automatically and across the visual field, a target that is distinct from its neighbors in its preattentive representation in the brain should "pop out" of the display. The proverbial needle in a haystack is hard to find because it shares properties of length, thickness and orientation with the hay in which it is hidden. A red poppy in a haystack is a much easier target; its unique color and shape are detected automatically.

We find that if a target differs from the distractors in some simple property, such as orientation or color or curvature, the target is detected about equally fast in an array of 30 items and in an array of three items. Such targets pop out of the display, so that the time it takes to find them is independent of the number of distractors. This independence holds true even when subjects are not told what the unique property of the target will be. The subjects take slightly longer overall, but the number of distractors still has little or no effect.

On the other hand, we find that if a target is characterized only by a conjunction of properties (for example, a red O among red N's and green O's), or if it is defined only by its particular combination of components (for example, an R among P's and Q's that together incorporate all the parts of the R), the time taken to find the target or to decide that the target is not present increases linearly with the number of distractors. It is as if the subjects who are placed in these circumstances are forced to focus attention in turn on each item in the display in order to determine how the item's properties or parts are conjoined. In a positive trial (a trial in which a target is present) the search ends when the target is found; on the average, therefore, it ends after half of the distractors have been examined. In a negative trial (in which no target is present) all the distractors have to be checked. As distractors are added to the displays, the search time in positive trials therefore increases at half the rate of the search time in negative trials.

The difference between a search for simple features and a search for conjunctions of features could have implications in industrial settings. Quality-control inspectors might, for example, take more time to check manufactured items if the possible errors in manufacture are characterized by faulty combinations of properties than they do if the errors always result in a salient change in a single property. Similarly, each of the symbols representing, say, the destinations for baggage handled at airline terminals should be characterized by a unique combination of properties.

In a further series of experiments on visual-search tasks, we explored the effect of exchanging the target and the distractors. That is, we required subjects to find a target distinguished by the fact that it *lacks* a feature present in all the distractors. For example, we employed displays consisting of O's and Q's, so that the difference between the target and the distractors is that one is simply a circle whereas the other

is a circle intersected by a line segment [*see illustration on page 120*]. We found a remarkable difference in the search time depending on whether the target was the Q and had the line or was the O and lacked the line. When the target had the line, the search time was independent of the number of distractors. Evidently the target popped out of the display. When the target lacked the line, the search time increased linearly with the number of distractors. Evidently the items in the display were being subjected to a serial search.

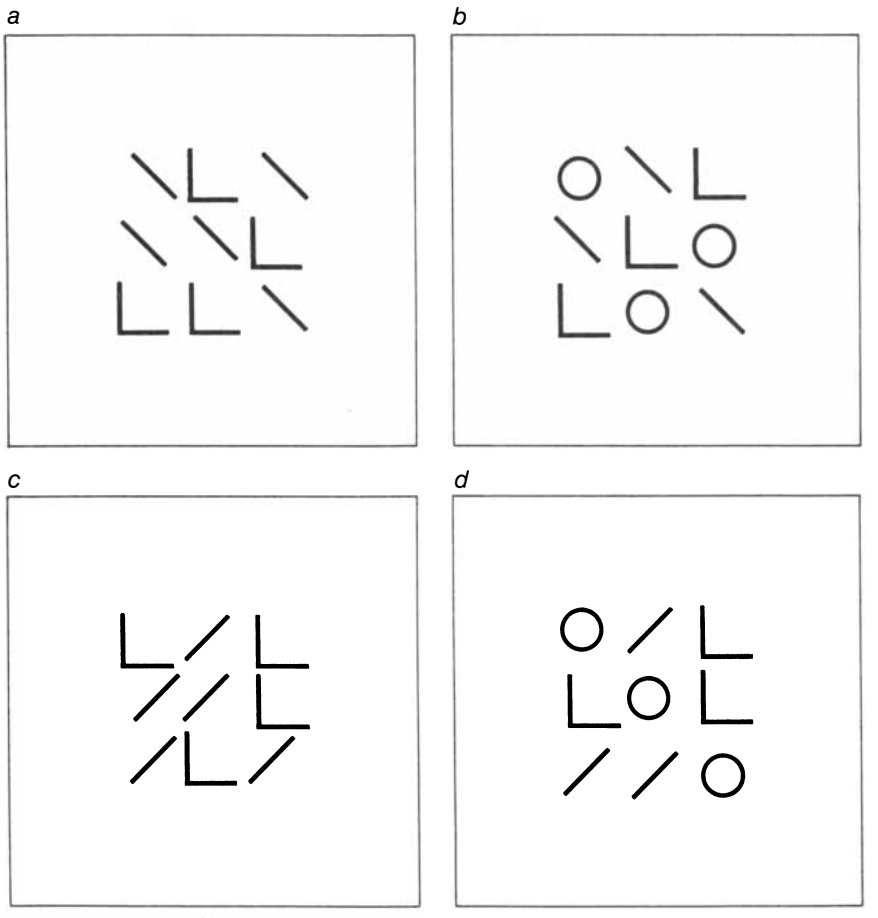
The result goes against one's intuitions. After all, each case involves the same discrimination between the same two stimuli: O's and Q's. The result is consistent, however, with the idea that a pooled neural signal early in visual processing conveys the presence but not the absence of a distinctive feature. In other words, early vision extracts simple properties, and each type of property triggers activity in populations of specialized detectors. A target with a unique property is detected in the midst of distractor items simply by a check on whether the relevant detectors are active. Conversely, a target lacking a property that is present in the distractors arouses only slightly less activity than a display consisting exclusively of distractors. We propose, therefore, that early vision sets up a number of what might be called feature maps. They are not necessarily to be equated with the specialized visual areas that are mapped by physiologists, although the correspondence is suggestive.

We have exploited visual-search tasks to test a wide range of candidate features we thought might pop out of displays and so reveal themselves as primitives: basic elements in the language of early vision. The candidates fell into a number of categories: quantitative properties such as length or number; properties of single lines such as orientation or curvature; properties of line arrangements; topological and relational properties such as the connectedness of lines, the presence of the free ends of lines or the ratio of the height to the width of a shape.

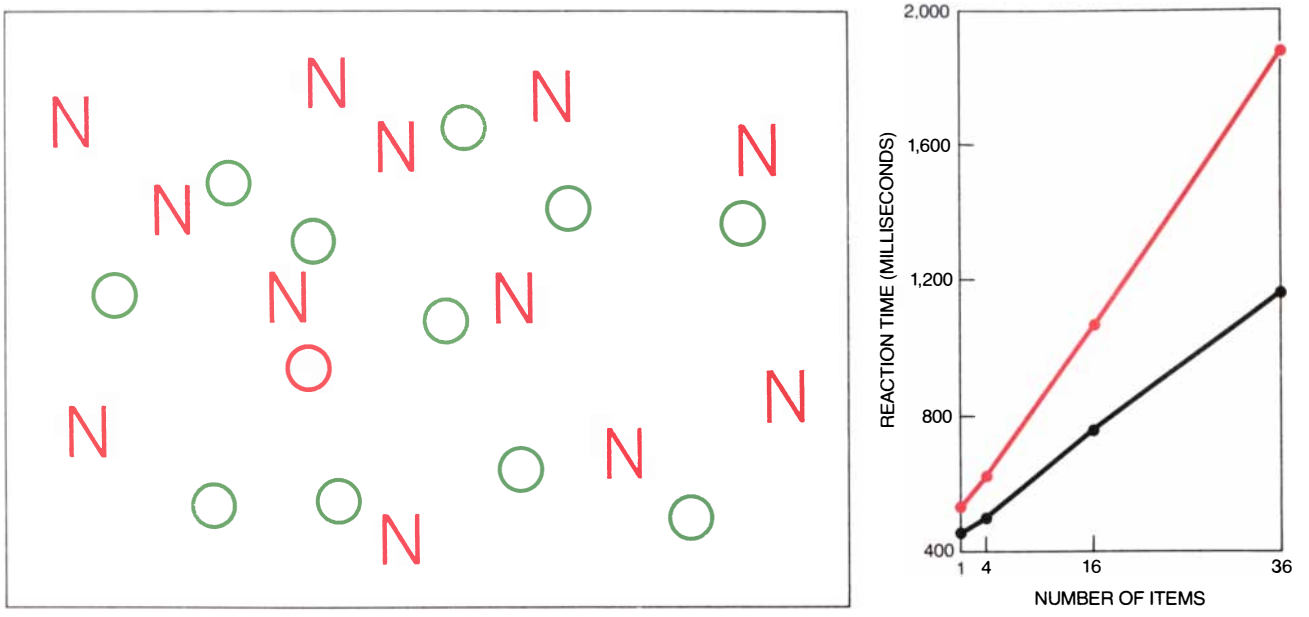
Among the quantitative candidates, my colleagues and I found that some targets popped out when their discriminability was great. In particular, the more extreme targets—the longer lines, the darker grays, the pairs of lines (when the distractors were single lines)—were easier to detect. This suggests that the visual system responds positively to "more" in these quantitative properties and that "less" is coded by default. For example, the neural activity signaling line length might increase with increasing length (up to some maximum), so that a longer target is detected against the lower level of background activity produced by short distractors. In contrast, a shorter target, with its concomitant lower rate of firing, is likely to be swamped by the greater activity produced by the longer distractors. Psychophysicists have known for more than a century that the ability to distinguish differences in intensity grows more acute with decreasing background intensity. We suggest that the same phenomenon, which is known as Weber's law, could account for our findings concerning the quantitative features.

Our tests of two simple properties of lines, orientation and curvature, yielded some surprises. In both cases we found pop-out for one target, a tilted line among vertical distractors and a curved line among straight lines, but not for the converse target, a vertical line among tilted distractors and a straight line among curves. These findings suggest that early vision encodes tilt and curvature but not verticality or straightness. That is, the vertical targets and the straight targets appear to lack a feature the distractors possess, as if they represent null values on their respective dimensions. If our interpretation is correct, it implies that in early vision tilt and curvature are represented relationally, as deviations from a standard or norm that itself is not positively signaled.

A similar conclusion emerged for the property of closure. We asked subjects to search for complete circles in the midst of circles with gaps and for circles with gaps among complete circles. Again we found a striking asymmetry, this time suggesting that the gap is preattentively detectable but that closure is not—or rather that it becomes preattentively detectable only when the distractors have very large



ILLUSORY TRIANGLES constitute a test of what features must be available to support the perception of triangles. Subjects seldom reported seeing a triangle when they were briefly exposed to displays consisting of the line segments that make up a triangle (*a*). They saw triangles far more often when the displays also included closed stimuli, that is, shapes that enclose a space, in this case O's (*b*). Evidently closure is a feature analyzed in early visual processing. This conclusion was supported by showing displays that lack the diagonal line to make a triangle (*c, d*). Subjects seldom saw triangles in such displays.



SEARCHING A DISPLAY for a target item defined by a conjunction of features proves to take longer the more "distractor" items the display contains. Here the target is a red O; the distractors are green O's and red N's. The target therefore differs from the distractors in its conjunction of shape and color. On trials when no target was present, the average search time increased by some 40 milliseconds for each distractor added to the display. On trials when a target was present (so that on the average subjects would have to search half of the items before finding the target), the search time increased at about half that rate. The implication is that a target characterized by a conjunction of properties requires subjects who are searching for the target to focus attention on each displayed item in turn. If a target has a unique color or shape, the number of distractors does not influence search time.

gaps (that is, when they are quite open shapes like semicircles). In other words, closure is preattentively detectable, but only when the distractors do not share it to any significant degree. On the other hand, gaps (or the line ends that gaps create) are found equally easily whatever their size (unless they are too small for a subject, employing peripheral vision, to see).

Finally, we found no evidence that any property of line arrangements is preattentively detectable. We tested intersections, junctions, convergent lines and parallel lines. In every case we found that search time increases with an increasing number of distractors. The targets become salient and obvious only when the subject's attention is directed to them; they do not emerge automatically when that attention is disseminated throughout the display.

In sum, it seems that only a small number of features are extracted early in visual processing. They include color, size, contrast, tilt, curvature and line ends. Research by other investigators shows that movement and differences in stereoscopic depth are also extracted automatically in early vision. In general the building blocks of vision appear to be simple properties that characterize local elements, such as points or lines, but not the relations among them. Closure appears to be the most complex property that pops out preattentively. Finally, our findings suggest that several preattentive properties are coded as values of deviation from a null, or reference, value.

Up to this point I have concentrated on the initial, preattentive stages of vision. I turn now to the later stages. In particular I turn to the evidence that focused attention is required for conjoining the features at a given location in a scene and for establishing structured representations of objects and their relations.

One line of evidence suggesting that conjunctions require attention emerges from experiments in which we asked subjects to identify a target in a display and say where it was positioned. In one type of display only a simple feature distinguished the target from the distractors. For example, the target was a red H in the midst of red O's and blue X's or an orange X among red O's and blue X's. In other displays the target differed only in the way its features were conjoined. For example, it was a blue O or a red X among red O's and blue X's.

We were particularly interested in the cases in which a subject identified the target correctly but gave it the wrong location. As we expected, the subjects could sometimes identify a simple target, say a target distinguished merely by its color, but get its location wrong. Conjunction targets were different: the correct identification was completely dependent on the correct localization. It does indeed seem that attention must be focused on a location in order to combine the features it contains.

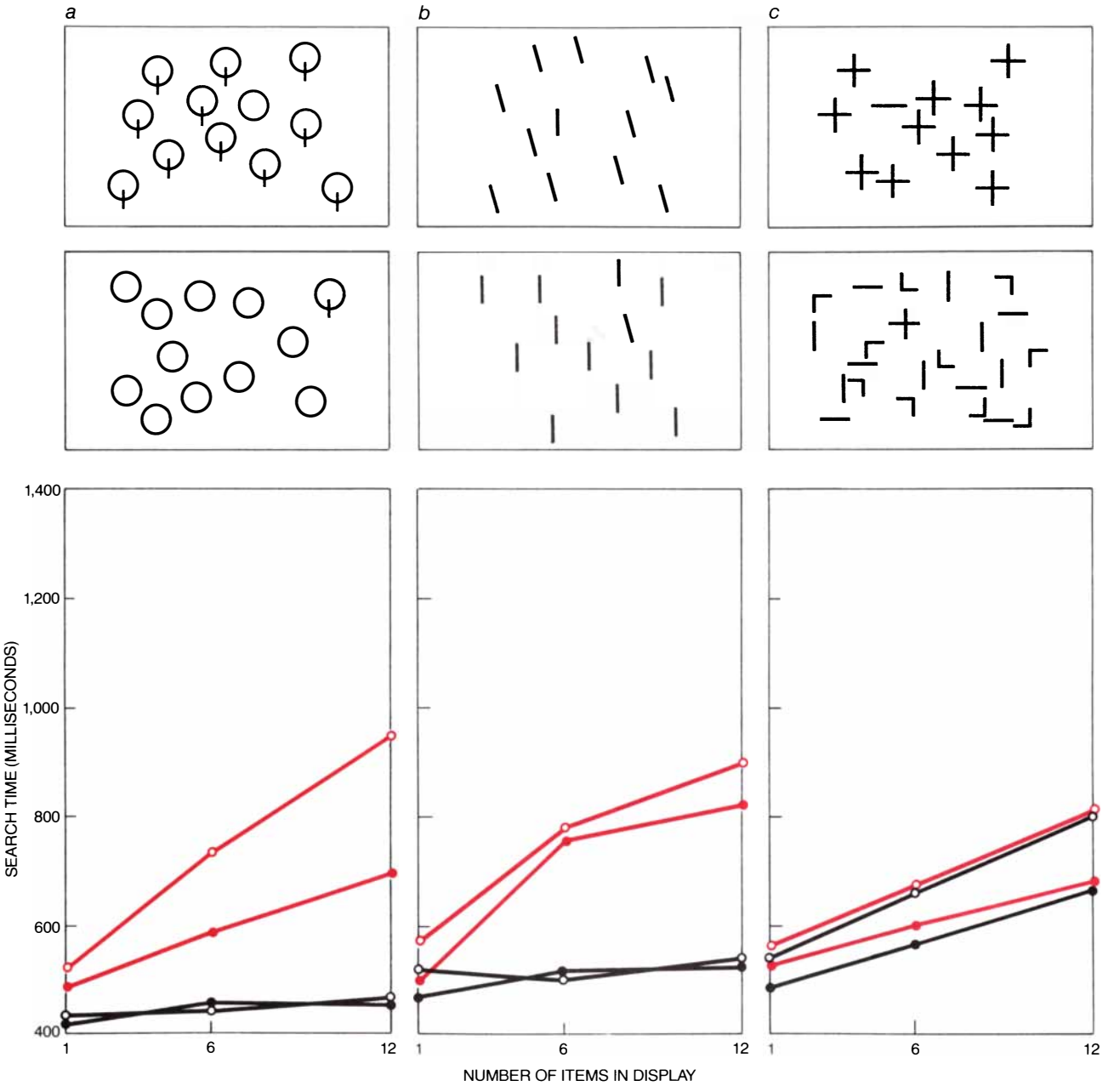
In a natural scene, of course, many conjunctions of features are ruled out by prior knowledge. You seldom come across blue bananas or furry eggs. Preattentive visual processing might be called "bottom up," in that it happens automatically, without any recourse to such knowledge. Specifically, it happens without recourse to "top down" constraints. One might hypothesize that conjunction illusions in everyday life are prevented when they conflict with top-down expectations. There are many demonstrations that we do use our knowledge of the world to speed up perception and to make it more accurate. For example, Irving Biederman of the State University of New York at Buffalo asked subjects to find a target object such as a bicycle in a photograph of a natural scene or in a jumbled image in which different areas had been randomly interchanged. The subjects did better when the bicycle could be found in a natural context [*see illustration on page 115*].

In order to explore the role of prior knowledge in the conjoining of properties, Deborah Butler and I did a further study of illusory conjunctions.

We showed subjects a set of three colored objects flanked on each side by a digit [*see top illustration on page 122*]. Then, some 200 milliseconds later, we showed them a pointer, which was accompanied by a random checkerboard in order to wipe out any visual persistence from the initial display. We asked the subjects to attend to the two digits and report them, and then to say which object the pointer had designated. The sequence was too brief to allow the subjects to focus their attention on all three objects.

The crucial aspect of the experiment lay in the labels we gave the objects. We told one group of subjects that the display would consist of "an orange carrot, a blue lake and a black tire." Occasional objects (one in four) were shown in the wrong color to ensure that the subjects could not just name the color they would know in advance ought to be associated with a given shape. For another group of subjects the same display was described as "an orange triangle, a blue ellipse and a black ring."

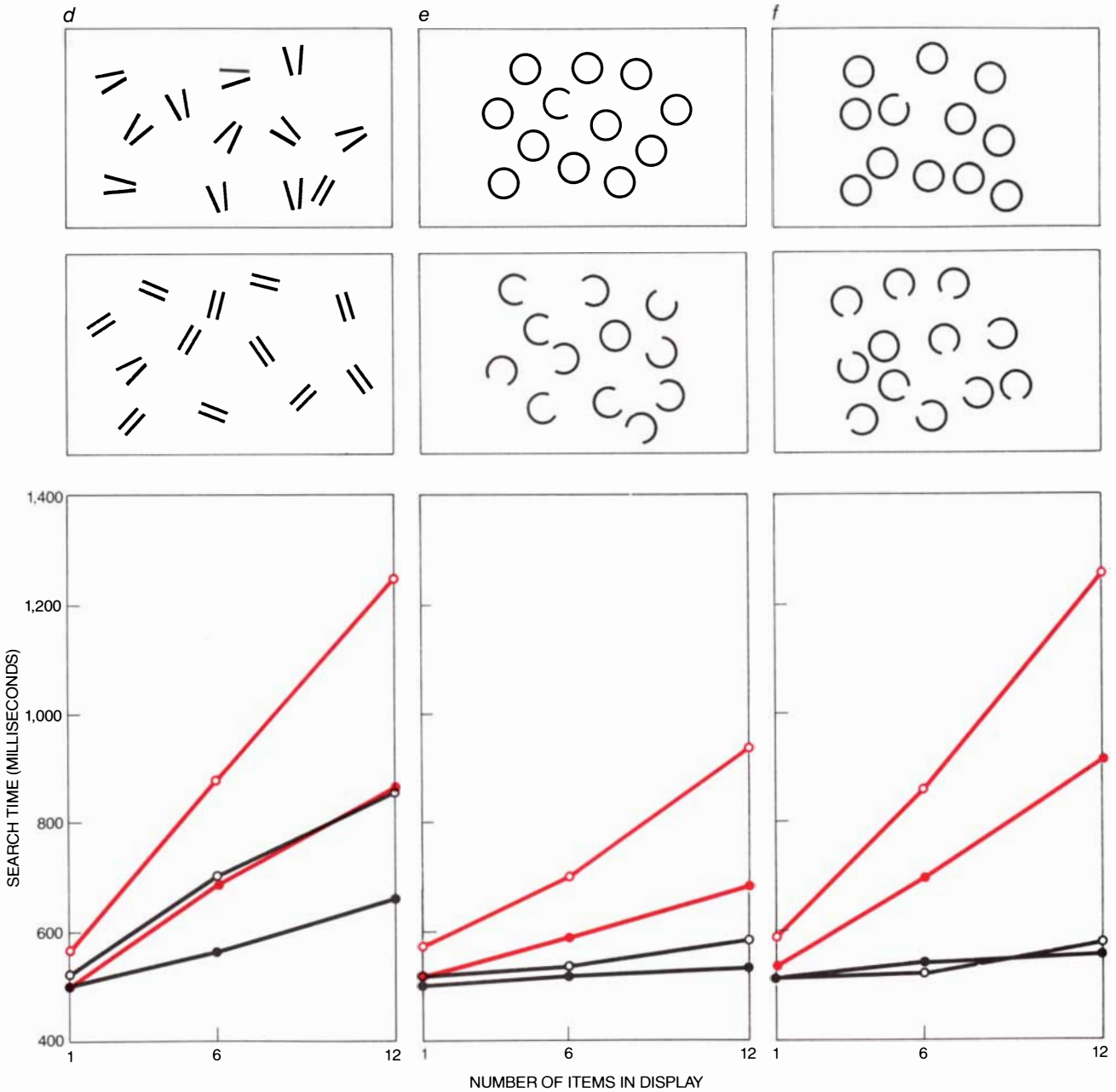
The results were significant. The group given arbitrary pairings of col-



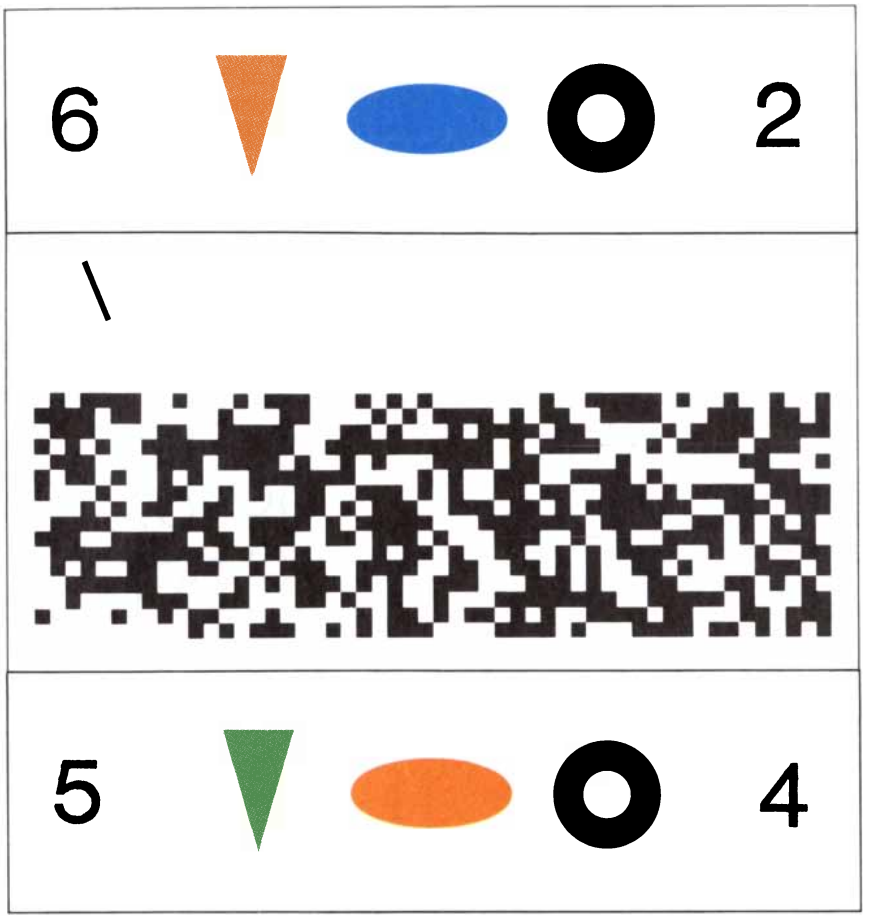
PRESENCE OR ABSENCE of a feature can have remarkably different effects on the time it takes to find a target in the midst of distractors. In one experiment (*a*) the target was a circle intersected by a vertical line segment or a circle without that feature. The search time for the intersected circle (*black*) proved to be largely independent of the number of items in the display, suggesting that the feature popped out. The search time for the plain circle (*color*) increased steeply as distractors were added, suggesting that a serial search of the display was being made. A second experiment (*b*) required subjects to search for a vertical line (*color*) or a tilted line (*black*). The tilted line could be found much faster; evidently only the tilted line popped out of the displays. A

ors and shapes reported many illusory conjunctions: 29 percent of their responses represented illusory recombinations of colors and shapes from the display, whereas 13 percent were reports of colors or shapes not present in the display. In contrast, the group expecting familiar objects saw rather few illusory conjunctions: they wrongly recombined colors and shapes only 5 percent more often than they reported colors and shapes not present in the display.

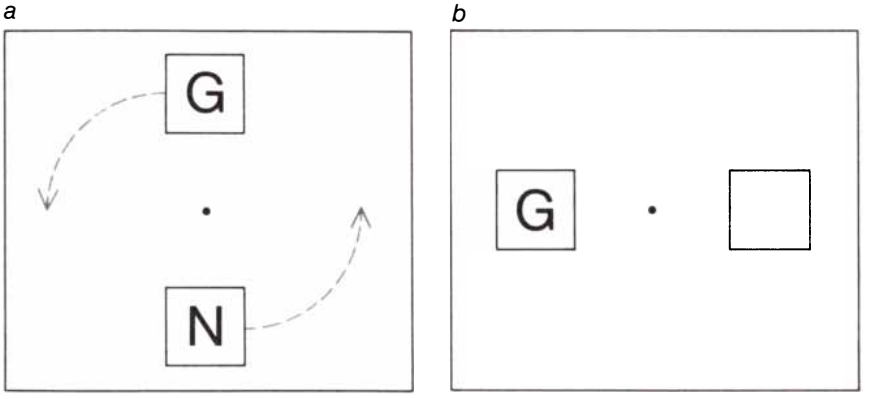
We occasionally gave a third group of subjects the wrong combinations when they were expecting most objects to be in their natural colors. To our surprise we found no evidence that subjects generated illusory conjunctions to fit their expectations. For example, they were no more likely to see the triangle (the "carrot") as orange when another object in the display was orange than they were when no orange was present. There seem to be two implications: prior knowledge and expectations do indeed help one to use attention efficiently in conjoining features, but prior knowledge and expectations seem not to induce illusory exchanges



third experiment (*c*) tested an isolated line segment (*color*) or intersecting lines in the form of a plus sign (*black*). Evidently neither popped out. A fourth experiment (*d*) tested parallel lines (*color*) or converging lines (*black*). Again neither popped out. A fifth experiment (*e*) tested closure with complete circles (*color*) or circles with a gap of a fourth of their circumference (*black*). A sixth experiment (*f*), again testing closure, had complete circles (*color*) or circles with smaller gaps (*black*). The size of the gap seemed to make no difference: the incomplete circle popped out. On the other hand, a complete circle became harder to find as the size of the gaps in distractors was reduced. Open dots represent data from trials in which the display included only distractors.



EFFECT OF EXPECTATIONS on the perception of conjunctions of features turns out to be complex. Subjects were shown three colored shapes flanked on each side by a distractor, specifically a digit (*top*). The display was followed by a masking field and a pointer (*middle*) indicating the prior location of the shape the subject was called on to report. The subjects made many mistakes in associating colors with shapes when they expected arbitrary pairings of colors and shapes (an orange triangle, a blue ellipse and a black ring). Not surprisingly, they made fewer mistakes when they were expecting pictures of familiar objects (a carrot, a lake and a tire). Some displays (*bottom*) showed unexpected combinations when the subjects thought they would see natural ones. Yet the subjects were no more likely to erroneously report, say, an orange carrot if orange was present elsewhere in the display than if orange was absent. These latter trials imply that illusory conjunctions are formed at a stage of processing that is not affected by prior knowledge.



INTEGRATION OF SENSORY INFORMATION into what amounts to a file on each perceptual object was tested by the motion of frames. In each trial two frames appeared; then two letters were briefly flashed in the frames (*a*). The frames moved to new locations, and a letter appeared in one of the two (*b*). The subject's task was to name the final letter as quickly as possible. If the final letter matched the initial letter and appeared in the same frame, the naming was faster than if the letter had appeared in the other frame or differed from the initial letter. The implication is that it takes more time to create or update a file on an object than it does simply to perceive the same object a second time.

of features to make abnormal objects normal again. Thus illusory conjunctions seem to arise at a stage of visual processing that precedes semantic access to knowledge of familiar objects. The conjunctions seem to be generated preattentively from the sensory data, bottom-up, and not to be influenced by top-down constraints.

How are objects perceived once attention has been focused on them and the correct set of properties has been selected from those present in the scene? In particular, how does one generate and maintain an object's perceptual unity even when objects move and change? Imagine a bird perched on a branch, seen from a particular angle and in a particular illumination. Now watch its shape, its size and its color all change as it preens itself, opens its wings and flies away. In spite of these major transformations in virtually all its properties, the bird retains its perceptual integrity: it remains the same single object.

Daniel Kahneman of the University of California at Berkeley and I have suggested that object perception is mediated not only by recognition, or matching to a stored label or description, but also by the construction of a temporary representation that is specific to the object's current appearance and is constantly updated as the object changes. We have drawn an analogy to a file in which all the perceptual information about a particular object is entered, just as the police might open a file on a particular crime, in which they collect all the information about the crime as the information accrues. The perceptual continuity of an object would then depend on its current manifestation being allocated to the same file as its earlier appearances. Such allocation is possible if the object remains stationary or if it changes location within constraints that allow the perceptual system to keep track of which file it should belong to.

In order to test this idea we joined with Brian Gibbs in devising a letter-naming task [*see bottom illustration at left*]. Two letters were briefly flashed in the centers of two frames. The empty frames then moved to new locations. Next another letter appeared in one of the two frames. We devised the display so that the temporal and spatial separations between the priming letter and the final letter were always the same; the only thing that differed was the motion of the frames. The subjects' task was to name the final letter as quickly as possible.

We knew that the prior exposure to a given letter should normally lessen the time it takes to identify the same

Versatile laser devices that can pinpoint targets for laser-homing weapons, conventional artillery, and naval gunfire are being delivered to the U.S. Marine Corps. Modular Universal Laser Equipment (MULE) can designate targets for all laser-guided weapons, including Laser Maverick, Hellfire, laser-guided bombs, and cannon-launched laser-guided projectiles. MULE is a portable tripod-mounted unit that consists of three modules: a laser designator/rangefinder, an instrument that finds true north, and a stabilized tracking tripod. The designator/rangefinder can be detached from the tripod and aimed by hand. The tripod displays range, azimuth, and angle of elevation of targets. MULE can combine azimuth, elevation, and range information into a digital message that can be sent to an automatic tactical fire control center. MULE is in production at Hughes Aircraft Company.

NATO will upgrade its air defense network with eight long-range radars for four of its member nations. The new radars are NATO versions of the HR-3000, a new generation derivative of the Hughes Air Defense Radar (HADR) operating in West Germany, Malaysia, and Norway. The radar is fully transportable and can be set up and torn down in hours. It also has better electronic counter-counter measures, improved capability for rejecting clutter, and a faster rotating antenna to accommodate NATO's requirement for a higher data rate. The radars will be installed in Turkey, Greece, and Italy, where they will be integrated into the Hughes-developed NATO Air Defense Ground Environment (NADGE) system, and in Portugal, where one will be integrated into a national air defense system.

Over 100,000 TV channels are now being carried to cable television subscribers in the U.S. by means of Hughes' AML microwave systems. AML (Amplitude Modulated Link) was developed as a way to deliver multichannel television programming to cable TV hub sites, much as a trunk cable does. An AML system can carry up to 80 TV channels simultaneously. At least half of the nation's cable subscribers in over 1,500 communities receive TV programming by this technique. AML microwave equipment is used in more than 500 cable TV systems in the U.S., Canada, Mexico, Belgium, Switzerland, Austria, Denmark, Finland, and Argentina.

An AMRAAM air-to-air missile can be fully tested in only one minute with a sophisticated test station. A typical station consists of 17 bays of state-of-the-art computer-controlled instrumentation and is designed for use in the development lab, military depot, or factory. Due to the thermal time limitation of the missile hardware, it examines all of the missile's functions in about 60 seconds—digital, analog, radio frequency, telemetry, and built-in-test capabilities. The central computer saves all information and analyzes it after the test is completed. The test stations are part of a full-scale development contract the U.S. Air Force awarded to Hughes for the Advanced Medium-Range Air-to-Air Missile. Each station can be expanded into a diagnostic station to pinpoint problem areas down to the smallest replaceable assembly and tell which repair should be made.

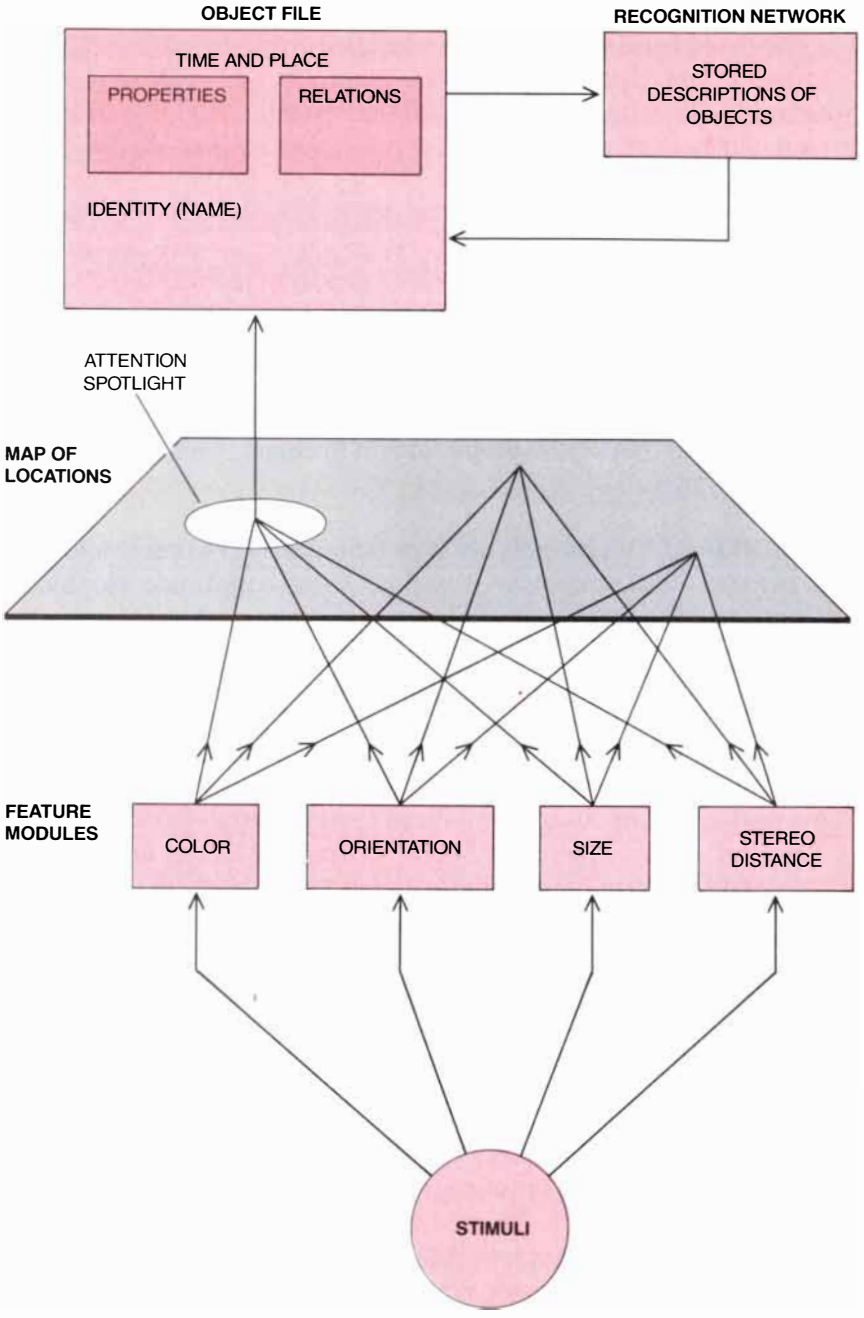
Support Systems in Southern California designs, develops, and manufactures some of the most sophisticated training simulators and a wide array of automatic and manual test systems. In addition, field engineering and technical support of a wide range of electronic systems keep Hughes' systems operating at top efficiency worldwide. Opportunities are available for a variety of engineers qualified by degree or extensive work experience. They include systems engineers, radar engineers, and software and hardware design engineers. Please send your resume to Lowell Anderson, Professional Employment, Dept S2, Hughes Aircraft Company, P.O. Box 9399, Long Beach, CA 90801-0463. 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

letter on a subsequent appearance; the effect is known as priming. The question that interested us was whether priming would occur only in particular circumstances. We argued that if the final letter is the same as the priming letter and appears in the same frame as the priming letter, the two should be seen as belonging to the same object; in this case we could think of the perceptual task as simply re-viewing the original object in its shifted position. If, on the other hand, a new letter appears in the same frame, the object file should have to be updated, perhaps increasing the time it takes for subjects to become aware of the letter and name it.



HYPOTHETICAL MODEL of the early stages in visual perception emerges from the author's experiments. The model proposes that early vision encodes some simple and useful properties of a scene in a number of feature maps, which may preserve the spatial relations of the visual world but do not themselves make spatial information available to subsequent processing stages. Instead focused attention (employing a master map of locations) selects and integrates the features present at particular locations. At later stages the integrated information serves to create and update files on perceptual objects. In turn the file contents are compared with descriptions stored in a recognition network. The network incorporates the attributes, behavior, names and significance of familiar objects.

Actually the priming was found to be object-specific: subjects named the final letter some 30 milliseconds faster if the same letter had appeared previously in the same frame. They showed no such benefit if the same letter had appeared previously in the other frame. The result is consistent with the hypothesis that the later stages of visual perception integrate information from the early, feature-sensitive stages in temporary object-specific representations.

The overall scheme I propose for visual processing can be put in the form of a model [*see illustration on opposite page*]. The visual system begins by coding a certain number of simple and useful properties in what can be considered a stack of maps. In the brain such maps ordinarily preserve the spatial relations of the visual world itself. Nevertheless, the spatial information they contain may not be directly available to the subsequent stages of visual processing. Instead the presence of each feature may be signaled without a specification of *where* it is.

In the subsequent stages focused attention acts. In particular, focused attention is taken to operate by means of a master map of locations, in which the presence of discontinuities in intensity or color is registered without specification of what the discontinuities are. Attention makes use of this master map, simultaneously selecting, by means of links to the separate feature maps, all the features that currently are present in a selected location. These are entered into a temporary object representation, or file.

Finally, the model posits that the integrated information about the properties and structural relations in each object file is compared with stored descriptions in a "recognition network." The network specifies the critical attributes of cats, trees, bacon and eggs, one's grandmothers and all the other familiar perceptual objects, allowing access to their names, their likely behavior and their current significance. I assume that conscious awareness depends on the object files and on the information they contain. It depends, in other words, on representations that collect information about particular objects, both from the analyses of sensory features and from the recognition network, and continually update the information. If a significant discontinuity in space or time occurs, the original file on an object may be canceled: it ceases to be a source of perceptual experience. As for the object, it disappears and is replaced by a new object with its own new temporary file, ready to begin a new perceptual history.

Galileo and the Specter of Bruno

The two men are often honored as martyrs to science, but for Bruno astronomy was a vehicle for politics and theology. Galileo was tried partly because his aims were mistakenly identified with those of Bruno

by Lawrence S. Lerner and Edward A. Gosselin

In 1633 Galileo Galilei was brought before the Roman Inquisition and charged with teaching that the earth moves. The ensuing trial engaged the full authority of the Roman Catholic Church in a scientific and theological dispute with the most important scientific figure of the day. As is well known, Galileo was forced to recant his scientific convictions publicly on the grounds that he had defended the Copernican view of the universe, a position condemned as "false and opposed to Holy Scripture" in 1616. The incident is now universally regarded as a critical event in the birth of modern science.

In 1600, well within living memory at the time of Galileo's trial, Giordano Bruno had also met his fate at the hands of the Roman Inquisition. Bruno too was a celebrated Copernican, but his penalty was much severer: he was burned at the stake. The careers and persecutions of Galileo and Bruno have been linked in two kinds of myth, which tend to confuse the two figures in curiously opposite ways. According to the first myth, both men challenged an ignorant and obscurantist Catholic church in a modern spirit of freedom, and both were martyred to their cause. In this context Bruno was seen as a kind of proto-Galileo, or perhaps a Galileo manqué. He was a proto-Galileo simply because he too espoused the Copernican system and suffered an extreme version of Galileo's fate. He was a Galileo manqué in that, lacking Galileo's scientific insight and genius, he resorted to speculative mysticism. There is a kernel of truth to this myth, as there is to many others, but the historical record yields a much richer and more elaborate tale.

The second myth is almost the converse of the first: that Galileo was a kind of resurrected Bruno in that he, like Bruno, was perceived as a religious and political revolutionary. The second myth must not simply be dismissed as wrongheaded. Indeed, our investigations have convinced us that an understanding of its role in early 17th-century beliefs is essential to a more complete understanding of what might be called the climatological prehistory of the trial of Galileo. We shall argue that Galileo's troubles came about in part because his contemporaries distorted his novel way of thinking into the more traditional and better-understood categories articulated by Bruno.

The first myth, depicting Bruno as a failed Galileo, was popular in the 19th century and well into the 20th; its moral was congenial to the world view of the 19th-century liberal. The myth lends support to the appealing legend that when Galileo recanted his Copernican position, he was heard to mutter, "Eppur si muove" ("And yet it moves"). There is no evidence to support that incident, and the real Galileo is a poor fit in such a heroic role. He is grossly miscast as a Nathan Hale or a William Tell.

What about Bruno? How does he fit the first myth that he was a martyr to science? For simplicity we shall confine our discussion mainly to his first and most important work on the Copernican system, *La Cena de le ceneri* ("The Ash Wednesday Supper," or *Supper* for short). It was written in Italian in 1584. At the time Bruno was a member of the London household of the French ambassador to England.

If one regards the work as an exposition of the Copernican system together with some other scientific topics, one can only puzzle over its notoriety. It appears to be a compendium of nonsense—a disorganized display of gross error connected by incomprehensible passages. Bruno has the Copernican model of the solar system wrong. He demonstrates total ignorance of the most elementary ideas of geometry, let alone geometric optics. He throws in scraps of pseudoscientific argument, mostly garbled, and proceeds to high-flying speculations that seem disconnected from the preceding or subsequent arguments. Even the diagrams do not always correspond to the accompanying discussions in the text.

But there must be more to it, and that deeper significance must have been evident to Bruno's contemporaries. If Bruno had merely been a fool, he might have met with laughter and derision but not violent enmity and vehement official response. The Roman Inquisition would surely not have gone to the trouble it did to silence and repress him: he was imprisoned for eight years and subjected to innumerable hearings and interrogations before he was finally executed.

Bruno repeatedly makes it clear that the *Supper* is really not about the Co-

PORTRAIT OF GALILEO by the 17th-century painter Ottavio Leoni depicts Galileo at the age of about 60. The time was one of great intellectual hope for Galileo and his followers: about a year earlier, in 1623, Maffeo Cardinal Barberini, a fellow Florentine, scholar and lover of the arts, had been elected Pope Urban VIII. A tapestry in the Vatican pictures Barberini defending a doctoral dissertation involving the accepted Ptolemaic astronomy, but he had followed Galileo's investigations of the Copernican model of the solar system with keen interest. The two men held a series of long conversations in May and June of 1624. Nine years later their friendship had changed to enmity, and Urban forced Galileo to recant his scientific convictions. The trial of Galileo was a political expedient for the pope, in part because in the public consciousness Galileo had come to resemble Giordano Bruno, a theological and political revolutionary and an early proponent of Copernican ideas, who was convicted of heresy and burned at the stake in 1600.

pernican system at all: it is only peripherally a work on natural science and it is emphatically not to be taken literally. In accordance with the title, its central subject is the nature of the Eucharist: the Christian ceremony of Communion. Bruno praises Copernicus in a qualified way as the light before a new dawn. He heaps scorn, however, on the anonymous author of the preface to Copernicus' work *De revolutionibus orbium coelestium,* for Bruno clearly recognizes that the preface is spurious: in a fit of caution it suggests Copernican astronomy should be understood as a theory for calculation only and not as an account of physical reality. (Today scholars know it was written by Andreas Osiander, a Lutheran theologian who was one of Copernicus' students.)

GIORDANO BRUNO is depicted in this anonymous engraving made in the early 17th century, about two decades after Bruno was executed by the Roman Inquisition. It is the earliest rendering of Bruno known, but no source contemporaneous with Bruno's life is known; the likeness is probably spurious. Bruno is shown in the habit of the Dominican order to which he belonged, a connection he acknowledged to varying degrees at different times in his life. Bruno's Dominican name on the inscription is in its Latin form. (His given name was Filippo.) "Nolanus" refers to his birthplace, Nola, a suburb of Naples.

For Bruno the value of the Copernican system lies not in its astronomical details but instead in its scope as a poetic and metaphoric vehicle for much wider philosophical speculation. The Copernican replacement of the earth by the sun at the center of the solar system is for Bruno a symbolic restoration of what he calls the ancient true philosophy; according to him, it is to this philosophy one must turn in order to understand the true meaning of the Eucharist.

The "ancient true philosophy" was Bruno's brand of Hermetism, a mystic view based on Neoplatonic writings of the second and third centuries A.D. In Bruno's day it was widely believed the author of these writings was the semidivine Hermes Trismegistus, said to be a contemporary of Moses. The Hermetic philosophy had supposedly been handed down through Orpheus, Zoroaster, Pythagoras, Plato and the later Greek philosophers, and separately through the Judeo-Christian tradition. In the process the "ancient true philosophy" had become debased and corrupted, which for Bruno accounted for the lamentable state of worldly affairs.

According to the Brunonian Hermetic view, man shares in divinity and is therefore at least potentially in constant communion with God. The universal divine principle is extended to the entities that make up the macrocosmic universe as well as to man. Thus both stars and planets (Bruno does not make a distinction) are animate and endowed with souls; they can be operated on magically to effect social and political change. Rediscovered and refurbished by Bruno, the "ancient true philosophy" would unfetter the divine essence within all men. Its power could heal religious and political wounds and give birth to a new golden age.

It is important to understand that Bruno's adoption of natural science to foster broader theological, ethical, social and political purposes was entirely characteristic of the Renaissance world view. For the people of the Renaissance science was literally a branch of philosophy, often called on to illuminate or illustrate a nonscientific issue. Intelligent and well-educated people often saw explicit and highly anthropocentric parallels between

scientific knowledge and the other aspects of life. Bruno is typical of contemporary Hermetists in leaping to conclusions about the relation of human beings to God based on theories about the workings of the macrocosm—and vice versa.

As Bruno saw it, one inevitable consequence of a return to the "true philosophy" would be the recognition of the essential unity between the liberal Protestants in England and the liberal Catholics in France. The theme of unity is sounded throughout the *Supper;* for example, in the Ptolemaic system there is a fundamental distinction between the earth, which was considered stationary, and the planets and stars, which were supposed to move within concentric spheres above the earth. On the other hand, once the Copernican view is accepted, the moving earth and the moving stars enjoy a similar status. Indeed, Bruno populated an infinite universe with an infinite number of stars, all in motion and central to their surroundings in an equivalent way. In like manner, he proposed a framework in which the liberal Protestant and Catholic views on the Eucharist could be seen as essentially equivalent.

What were the implications of this view for contemporary politics? To maintain the balance of power, England had traditionally allied itself with either France or Spain, whichever was weaker. In the late 16th century Spain was strong and religiously orthodox. France was in the throes of a violent and protracted religious civil war, in which Spain supported the orthodox Guisard party against the Protestants. The moderate French king Henry III and his adherents were caught in the middle. Nevertheless, an alliance between the French moderates and England, which seemed desirable to both parties, was blocked by an official commitment of France to Catholicism. A bridge over religious differences would therefore have been of great value in fostering the alliance.

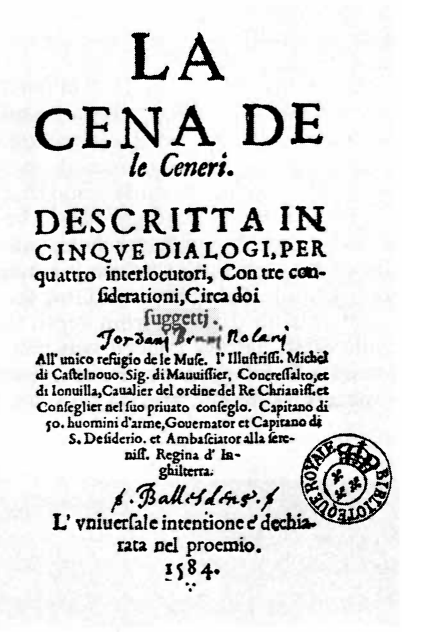
Bruno was convinced that a monarch inspired and enlightened by Hermetism could lead the world into a golden age under the guidance of the "true philosophy." His candidates for this role were successively Henry III, Elizabeth I of England and finally Henry IV of France. Indeed, two years after the publication of the *Supper* he confided to Guillaume Cotin, the librarian of the Abbey of St. Victor in Paris, that from 1582 to 1585 he had been a kind of intellectual ambassador from the liberal Catholics of Henry III to the liberal Protestant court of Elizabeth I. Bruno's later enthusiasm for the prospects of Henry IV was so great that in 1591 he imprudently interrupted his wanderings across Europe to return to Italy. His ultimate aim in returning seems to have been the conversion of the pope himself to Brunonian Hermetism.

Bruno's rash behavior almost immediately led to disaster. He was arrested on trumped-up charges by the Venetian Inquisition and then imprisoned. A year later he was moved to Rome, where he was confined for another seven years. In prison Bruno was no real threat to the Papacy. He had no following, no money and no influence. In the normal course of events he might simply have remained in prison until he was dead or forgotten. His execution seems to have been the pope's side of a minor political quid pro quo with the Spanish Hapsburgs.

In disposing of Bruno both the Papacy and Spain were announcing to all concerned that there would be no toleration of challenges either to religious orthodoxy or to the political status quo to which orthodoxy was closely linked, a connection we shall explore further. Nevertheless, Bruno's philosophical heritage survived in the early 17th century in a loose constellation of beliefs, held by an unorganized group of intellectuals called Rosicrucians. The aims and hopes of the Rosicrucians for a return of Pythagoreanism were nurtured in 1613 by the marriage of Frederick V, elector of the Rhenish Palatinate, to Elizabeth Stuart, daughter of James I of England. The marriage, mystically called the "chemical wedding of the Thames and the Rhine," was supposed to presage a return to the golden age of Elizabeth I.

In 1618 Frederick and Elizabeth were elected king and queen of Bohemia, from time to time a center of religious liberalism and toleration. Almost immediately they were deposed by the Hapsburg Holy Roman Emperor; the incident was the opening salvo of the Thirty Years' War, which renewed the political and religious strife of the late 16th century. In the years immediately following there was a "Rosicrucian scare": a strong and irrational fear of Rosicrucian subversion in Catholic strongholds.

Such was the prevailing political and religious climate when Galileo's popular exposition of his own Copernican views, *Dialogue on the Two Great World Systems,* was published in 1632. The publication had profound significance for people whose memories of the religious conflicts of the late 16th century were fresh and detailed. In a political and religious atmosphere


LA
CENA DE
le Ceneri.
DESCRITTA IN
CINQVE DIALOGI, PER
quattro interlocutori, Con tre consideratioui, Circa doi
suggetti.
Iordano Bruno Nolano
All' unico refugio de le Muse. l' Illustriss. Michel di Castelnouo. Sig. di Mauuissier, Concressalto, et di Ionuilla, Caualier del ordine del Re Chriauiss. et Conseglier nel suo priuato conseglo. Capitano di 50. huomini d'arme, Gouernator et Capitano di S. Desiderio. et Ambasciator alla serenissi. Regina d' Inghilterra.
L' vniuersale intentione e dechiarata nel proemio.
1584.

BRUNO'S STATEMENT of the Copernican picture of the universe is *La Cena de le ceneri* ("The Ash Wednesday Supper"); the title page from the book is shown. The work was written in Italian and published in England in 1584; an English translation by the authors of this article came out in 1977. The book is in the form of a debate centered on the Copernican system; Bruno's real objective was to extend Copernican ideas allegorically to theological and political speculation. The subtitle leads inexorably to the theme of unity emphasized throughout the work: "Described in five dialogues, by four interlocutors, with three reflections, on two subjects [dedicated] to the sole refuge of the Muses."

bearing a remarkable resemblance to that of the 1590's it was all too easy to confuse the aims of Galileo with those of Bruno. Hence the second and converse myth arises: in the public mind of 1633 Galileo became a revival, albeit an unwilling one, of the archetypal symbol of political and philosophical upheaval that Bruno had been three decades earlier.

We now know that Galileo's thinking was a radical departure from Renaissance thought. Indeed, it has often been maintained that Galileo's greatest contribution was his way of thinking about the physical universe. Unfortunately that was not necessarily clear to the great majority of his contemporaries. They were unable to understand not only his methods and his conclusions but also his aims and his intentions.

Compounding the lack of understanding was the fact that the late-Renaissance reader was predisposed to expect deliberately obscure writing.

Such obscurity was common, even when there was no external motive for it. In part its popularity is explained by a fondness for ornamentation and for elaborate elegance of expression. There was also a strong feeling that knowledge was for the initiate and that worthwhile understanding could be achieved only through much travail. The writing of Galileo's time is often full of multiple meanings, enabling the diligent reader to move from depth to depth. Such literary conceit was considered to be an appropriate reflection of the natural world, which yields progressively deeper secrets to the diligent and gifted. Needless to say, Bruno employed this approach extensively.

Galileo emphatically did not write in such a manner, but it is much easier to appreciate this fact today than it was in the early 17th century. At that time there were some people inclined to believe his writing also concealed multiple levels of meaning; in particular, his *Dialogue* could readily be seen as a sequel to Bruno's *Supper*.

What are the grounds on which such a view was based? First one can list what might be called external literary reasons. Earlier Galileo had written a work called *The Starry Messenger,* whereas in the *Supper* Bruno called the stars "messengers of God." Furthermore, both Bruno's *Supper* and Galileo's *Letter to the Grand Duchess* had expressed similar attitudes toward Holy Scripture. Both works maintained that the Bible often speaks according to the common understanding of the people, and in so doing it may actually say things about nature that are not literally true. The view was eccentric enough to have caused Galileo and Bruno to be closely associated. Finally, it is worth mentioning that both Galileo and Bruno called the Copernican theory Pythagorean, a name often used synonymously with the politically unsavory term Hermetic.

ROBERT CARDINAL BELLARMINE played an important role in deciding Bruno's fate in 1600, and he personally admonished Galileo in 1616 not to hold to or defend the Copernican doctrine that the earth moves. It was clear to Galileo from the interview, however, that Bellarmine did not exclude the study of the Copernican view as a scientific hypothesis. Bellarmine's death before Galileo's trial in 1633 left an ambiguity in the position of the church as it had previously been stated to Galileo. This painting of Bellarmine is in the Roman church of St. Ignatius, where Bellarmine's remains are also kept.

Second, both Bruno's *Supper* and Galileo's *Dialogue* are remarkably similar in form. For example, both dialogues introduce courtly gentlemen and dogmatic Aristotelians as interlocutors. In both dialogues there is extensive discussion of a thought experiment in mechanics: If a stone is dropped from the top of the mast of a moving ship, where will the stone land on the deck? Both works also give the same correct answer: The stone will land at the foot of the mast. No matter that the points made by the thought experiment in the two works are entirely different, or that Bruno's arguments are garbled when they are taken in a purely physical way. A suspicious mind not inclined to physics would find it easy to confuse Galileo's intentions with Bruno's. Furthermore, both dialogues are written in Italian rather than in Latin. The reasons for this linguistic departure from contemporary scholarly convention are now known to be quite different, but again a suspicious reader might think in both cases the vulgar tongue was adopted in order to stir debate and intellectual dissent on the widest scale possible.

There is another and more global similarity in the form of the two works that could easily lead to the wrong conclusion about Galileo's intentions. In both works the Copernican system is defended in a nonmathematical and nontechnical way, completely ignoring the very details that make the system interesting and workable for the practicing astronomer. We have already shown that Bruno exploits the Copernican vision as a hieroglyph for his religious and reunionist aims. Galileo's dialogue is a purely philosophical and scientific work, but it too does not really defend the Copernican system in its then current form. Instead Gali-

leo widens his discussion to judge between his own philosophy of nature and the Aristotelian one. The *Dialogue* is thus opened to the suspicion that its chief aim lies below the surface, as is demonstrably the case for the *Supper*.

The close link between politics and religion we mentioned above underlies a third major reason for confusing Galileo with Bruno. Copernicanism was widely associated with revolutionary political and religious thought. In 1599 Tommaso Campanella, an apostate Dominican monk, led a revolt against Hapsburg rule in Calabria, in the southern part of Italy. Campanella's aim was to reestablish the famous ancient Pythagorean city of Croton in Calabria, but his revolt was quickly crushed and he spent the next 26 years in prison. During those years he was far from idle. In 1600 he wrote his *City of the Sun,* a distinctly Hermetic and Copernican work (although it was not Copernican in an astronomical sense). Within a few years of his arrest his jailers lost interest in him, and he was left in relative peace to write, which he did voluminously.

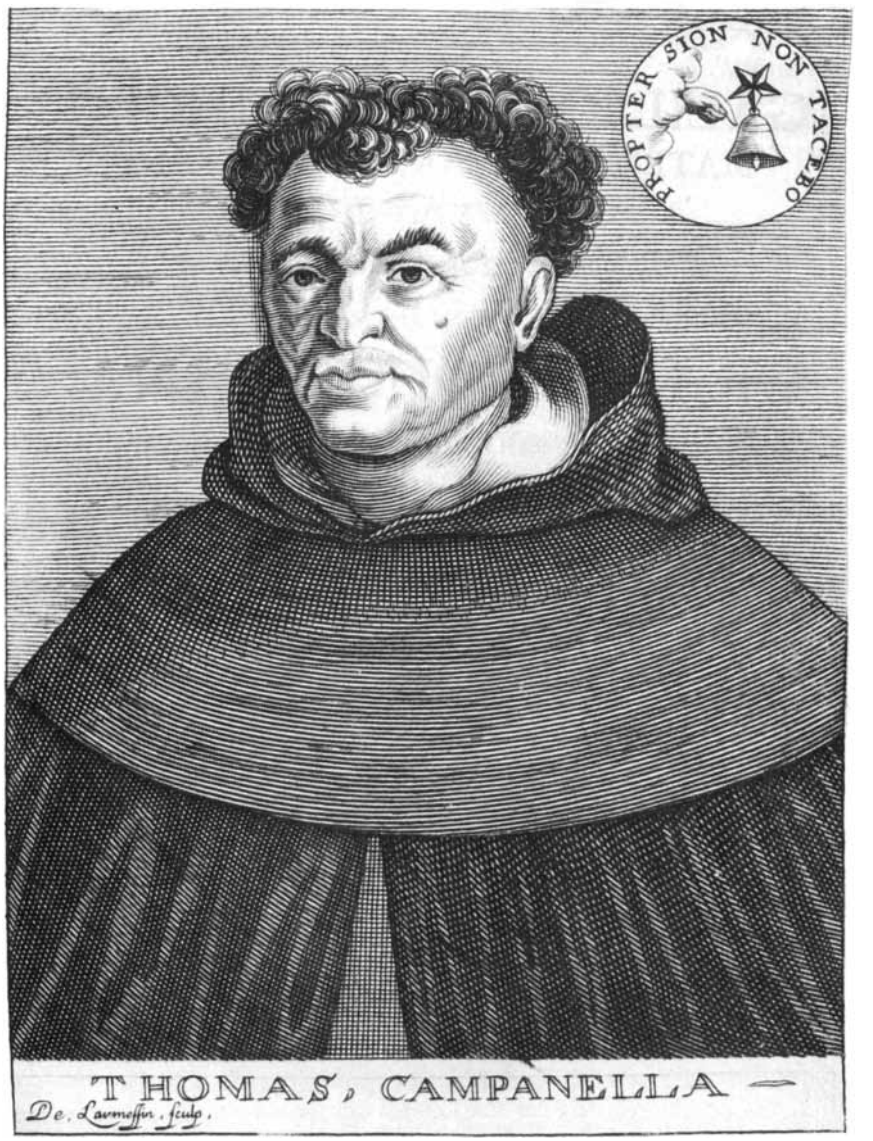
In 1616 Campanella learned that the Jesuit Robert Cardinal Bellarmine planned to admonish Galileo with respect to the dangerous theological implications of his scientific teachings. Campanella hastily wrote an *Apologia pro Galileo,* in which he attempted to show that Galileo's views were actually more rather than less in accordance with accepted theology than the Ptolemaic system was. The theological arguments presented by Campanella on this point bear strong similarities to those made by both Bruno and Galileo. To the possible later detriment of Galileo's case, the *Apologia* is the only work by Campanella in which Bruno is mentioned at all.

Campanella also wrote several letters to Galileo from prison. The letters make it clear that he viewed Galileo much as Bruno had viewed Copernicus: as a genius who heralded the dawn of the new truth without comprehending the philosophical significance of his discoveries. Such an attitude was almost certainly shared by others who were much less friendly to Galileo than Campanella was.

We strongly suspect that Galileo's close, if unwanted, association with Campanella was a major element in Galileo's eventual recantation. When Campanella was finally released from prison, he came into the pope's favor. In the late 1620's the Spanish grew increasingly angry with Urban VIII for following a partly secret and rather clumsy pro-French policy. Consequently they engaged in psychological warfare that played on Urban's belief in astrology: astrologers predicted the pope's death with the solar eclipse of June, 1628, and later with the eclipse of December, 1630. Campanella was known to be a magus, or sorcerer, and so Urban enlisted his aid in performing antieclipse magic. The magic was apparently successful, because Urban did not die.

For a few years Campanella's success admitted him to the confidence of the pope. It was necessary, however, to keep such magical operations a close secret. In 1600 the Jesuit Martin del Rio expressed an opinion of the church, in which many forms of magic were condemned. It would not do to let the public see the pope engaged in activity that was arguably heretical.

Campanella capitalized on the success of his antieclipse magic to put forward his Hermetic views, and to some degree he was successful. Indeed, at



TOMMASO CAMPANELLA, shown here in a 17th-century engraving by Nicolas de Larmessin the Elder, was an apostate Dominican monk who was imprisoned in 1599 for leading a revolt against the rule of the Hapsburgs in southern Italy. Tortured, starved and frozen while in prison, he nonetheless wrote voluminously about Hermetic and philosophically Copernican ideas. In 1616 he wrote his *Apologia pro Galileo* ("Apology for Galileo"), in which he argued that Galileo's views on astronomy were in accordance with orthodox Roman Catholic theology. Campanella later enjoyed the pope's confidence for a time, but he again fell from favor. He figures prominently in the confluence of events that finally brought Galileo to trial before the Inquisition. The motto *Propter sion non tacebo* recalls Campanella's zeal: "For the sake of the Heavenly City I will not keep still."

one point Urban authorized Campanella to found a college of missionaries, bearing the family name Barberini, which would convert the world to a "reformed," "natural" Catholicism. (Such adjectives were Hermetic watchwords.) Urban found Campanella's Hermetism congenial and was encouraged in his own purely political pro-French policy. He entered into an alliance with France. In the meantime, however, the French had given their support to the Protestant Swedish king, Gustavus Adolphus. By 1632 the king's military successes had been so sensational that both the political viability of the Holy Roman Empire and the religious viability of the Roman church throughout much of Germany were seriously threatened.

Thus the pro-French policy turned out to be too successful, and the growing hostility of the Hapsburgs could no longer be ignored. In such a political climate it is not surprising that Urban began to regard Campanella as a false prophet and all his teachings as dangerous. Since Campanella, Bruno, Hermetism and Copernicanism were all tangled together, it would have been a simple step in Urban's mind to link them with another eminent Copernican, Galileo. Given the proper background of fear and suspicion, this association and others could have inclined the pope and the Roman authorities to mistakenly regard Galileo as an antipapist, a religious liberal and perhaps a Hermetist.

DIALOGO
DI
GALILEO GALILEI LINCEO
MATEMATICO SOPRAORDINARIO
DELLO STVDIO DI PISA.
E Filoſofo, e Matematico primario del
SERENISSIMO
GR. DVCA DI TOSCANA.
Doue ne i congreſſi di quattro giornate ſi diſcorre
ſopra i due
MASSIMI SISTEMI DEL MONDO
TOLEMAICO, E COPERNICANO;
Proponendo indeterminatamente le ragioni Filoſofiche, e Naturali
tanto per l'vna, quanto per l'altra parte.
CON PRI VILEGI.
IN FIORENZA, Per Gio:Batiſta Landini MDCXXXII.
CON LICENZA DE' SVPERIORI.

TITLE PAGE from Galileo's *Dialogue on the Two Great World Systems* is decorated with a printer's colophon, or emblem, that depicts three fish swimming after one another in a circle. The unintended symbolism of the emblem caused a furor within the Roman church. The fish were assumed to be dolphins—albeit without justification to the modern eye—and the dolphins were immediately taken as a representation of Hermetic and Brunonian ideas condemned by the church. In the contemporaneous context of the Thirty Years' War the emblem was also interpreted as a symbol of philosophical and theological support for the French side. The motto of the emblem, *Grandior ut proles* ("Greater than the offspring"), was misinterpreted to imply that Galileo was extending the work of Bruno by harking back to a superior predecessor, perhaps Apollo, Pythagoras or Hermes Trismegistus. Such figures were associated with the officially discredited Hermetic philosophy.

There was a fourth reason for the Renaissance confusion of Galileo with Bruno, which seems peculiar to the modern mind. On the title page of Galileo's *Dialogue* his printer, Giovanni Batista Landini, had added a colophon, or emblem, which depicted three dolphins swimming after one another in a circle [*see illustration on opposite page*]. The emblem could be—and was—construed in an extremely damaging sense. In a letter to a friend of Galileo's in Florence, Galileo's Roman disciple Filippo Magalotti gives an account of an interview with Father Niccolò Riccardi. Riccardi was a Florentine who had risen high in the papal bureaucracy to become Master of the Sacred Palace, that is, the pope's theologian. Magalotti writes in part:

"On Monday morning...the most reverend Father...came to seek me.... Then he proceeded to disclose another reason for wishing to have...the *Dialogue*.... Under the seal of secrecy he told me that great offense had been taken at the emblem which was on the frontispiece.... On hearing this, I burst out laughing...and said I thought I could assure him that Signor Galileo was not the man to hide great mysteries under such puerilities and that he had said what he meant clearly enough. I declared that I believed I could affirm that the emblem was the printer's own. On hearing this, he appeared greatly relieved and told me that, if I could assure him that such was the case...the result would be most happy for the author."

The dolphin was a fairly common element in Renaissance emblematics, but the emblem in the *Dialogue* was not a standard cliché. The quest for its "real" meaning was one of the factors that stirred up the first unease about

the work in Rome. Riccardi was favorably inclined toward Galileo, and his agitation, together with his implication that the entire case against Galileo depended on the colophon, shows how important it must have been. It was not uncommon to convey arcane messages through symbols. The Roman authorities were led to suspect that Galileo and Landini had together contrived the emblem to hide a message, or perhaps to act as a key or a summary of the work that was to follow.

A suspicious nature could devise connotations of the emblem that were particularly dangerous to Galileo. First the dolphin was associated with the shrine of the god Apollo at Delphi. In Greek mythology Apollo was the father of Asclepius, who was one of the major figures of the Hermetic myths. Furthermore, in Homer's *Iliad* Apollo had been the leading divine supporter of the Trojans, and one of the Trojan survivors, Francus, was the legendary founder of the French royal line. To reinforce the speculation, the word "dauphin" is etymologically related to "dolphin." In the context of the Thirty Years' War such a linkage of France and Troy might be open to serious misinterpretation: as a passing of the mantle of Apollo—or in Christian terms the passing of the Holy Spirit that confers authority on the pope—from Rome to the rising collateral lineage of the Trojans in France.

All such speculation later evaporated. After an embarrassing delay Magalotti was finally able to give Riccardi another, earlier Landini book with the same colophon. By that time, however, the fat was in the fire: Riccardi was off in full cry over the theology of the *Letter to the Grand Duchess,* to whose existence in manuscript Magalotti had clumsily called attention. We have already mentioned some of the difficulties raised by the *Letter;* from then on Galileo's adversaries had more substantial issues to chew on.

We can now describe how the spurious association of Galileo with the Hermetic tradition meshed with contemporaneous political reality and gave a powerful initial impetus to the judicial machinery in which Galileo was caught. Urban VIII preferred to avoid the opposing extremes of religious orthodoxy and Counter-Reformation enthusiasm. His political and personal inclinations led him on a course clear of Spain and thus toward France. As we have mentioned, however, his actions had stirred growing anger among the Spaniards.

In 1631 Gaspare Cardinal Borgia, the Spanish ambassador to the Holy See, put increasingly heavy pressure on the pope for positive support of the Spanish cause. Urban recognized the danger of growing Spanish hostility, and he declared his readiness to try to dissuade France from an alliance with Sweden. In spite of this move, in March of 1632 Borgia openly attacked the pope in consistory, the meeting of the cardinals. Urban was furious, but he held his peace for fear of an open rupture with Philip IV of Spain.

To mollify Spain, Urban might have moved against France, but such a move would have risked an open break with the French church. The only course open to him was to make a nonsubstantive but strongly symbolic gesture. He could publicly sacrifice a person associated with the pro-French policies and the Hermetic philosophy from which he badly needed to dissociate himself. Campanella was the obvious choice. He was clearly connected not only with the pro-French policy but also with its underlying philosophy. Besides, he was expendable.

But Campanella knew too much. If strong and direct action were taken against him, the unedifying story of a pope who practiced heretical magic would certainly have come out. The next-best choice was Galileo. His association with Campanella in the public mind is underscored by the fact that the writer of the preliminary report of the committee investigating the *Dialogue* at first penned in Campanella's name instead of Galileo's and then had to erase it.

Given even these demanding circumstances, Urban might well have searched further for a scapegoat inasmuch as he had long held Galileo in high regard. Unfortunately, with singularly bad timing Galileo had managed to hurt Urban's feelings. The pope had argued that God could have accomplished his ends in infinitely many ways; as he put it to Galileo, "You must not necessitate God." Because Urban knew of Galileo's Copernican views, he had instructed Galileo not to make any definitive choice between the Ptolemaic and the Copernican systems. Galileo followed the instructions to the letter: the *Dialogue* ends with a passage disclaiming any choice between the two positions. The disclaimer, however, is clearly hypocritical, and Urban's views are put into the mouth of the dogmatic and dull-witted Simplicio. Consequently Urban was not unwilling to make a sacrifice of Galileo when one was needed.

Hence three strands are seen to converge. The political situation demanded a sacrifice. Galileo's writings, interpreted by men who were intelligent but out of touch with his real intent, made him a likely candidate. His supposed mocking of the pope deprived him of Urban's goodwill at a crucial moment. The Hermetic misinterpretation of his writings had been able to launch events that were later sustained by other forces.

One of the least convincing of the traditional explanations for Galileo's troubles has been that his trial went forward because of the pope's personal animosity. Urban's feelings alone would not have led him to set the immense machinery of the Inquisition against Galileo. The enormous disparity between the social and the political positions of Urban and Galileo would have made such a response to a personal affront entirely inappropriate. Our extension of the accepted explanation for the trial shows how Urban's strong reaction can be understood. Given the exigencies of the state and the fact that Galileo seemed to fill a political need, the momentary agreement between this need and Urban's personal pique made things go badly for Galileo.

One can also understand the ambiguities of Urban's later treatment of Galileo. Galileo was old and in poor health when he was summoned to Rome, but even when he lost his sight, Urban angrily refused to mitigate his conditions of house arrest. Yet Urban never tried to interfere with the continued lionization of Galileo, even by princes of the church. Nor did Urban make any attempt to interfere in any effective way with Galileo's freedom to publish, which he surely would have done if he had regarded Galileo as dangerous outside the context of the Spanish policy. Once Galileo had served the purpose of being a symbolic victim, Urban's further actions seemed to stem from mere personal anger rather than official enmity.

Galileo thus became a symbolic victim in an age that set great store by symbols, just as Bruno had before him. He was a victim of pan-European reputation and high visibility, and his exemplary punishment again taught the lesson of 1600: that anti-Hapsburg, pro-French, religiously conciliatory policies were heretical or at least close to it. A caricature of Galileo was concocted from the real man by his enemies and the circumstances of the moment, and so it was a resurrected Bruno who went before the Inquisition in 1633 to furnish a myth around which an existing moral could be woven.

THE AMATEUR SCIENTIST

The hyperscope and the pseudoscope aid experiments on three-dimensional vision

by Jearl Walker

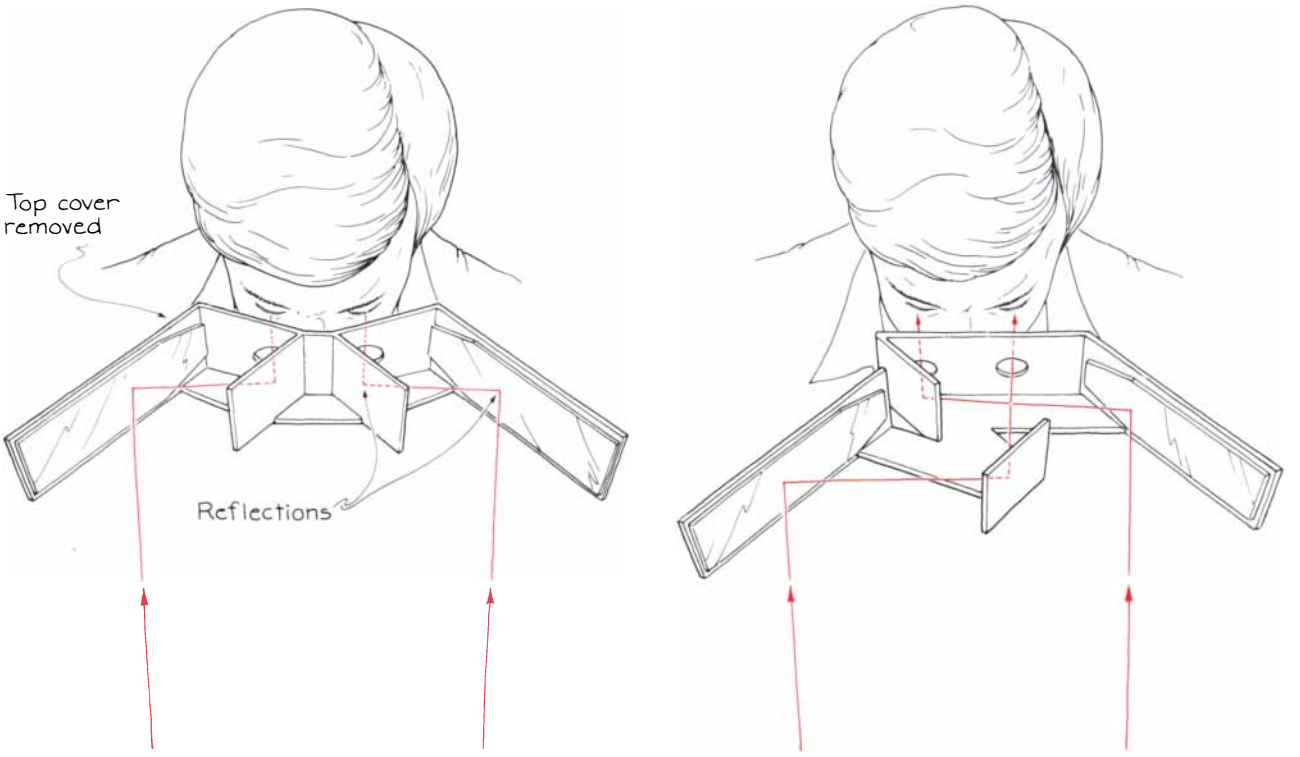
How is it that the visual system sees things three-dimensionally when the image on the retina is two-dimensional? The reason is that one interprets a variety of cues in the retinal images to create a perception of depth in a scene. Terry Pope of the University of Reading has devised two instruments that alter the cues so that he can do experiments on the perception of three dimensions.

The cues about distance and depth can be grouped into five categories: convergence, retinal disparity, accommodation, motion parallax and pictorial. Convergence involves the angle between the lines of sight from each eye when you look directly at an object. Retinal disparity involves the difference in the position of an image on the two retinas. Accommodation is a change in the shape of the eye's lens in order to focus an object onto the retina. Motion parallax is the relative motion of near and far objects through your field of view when you move or the objects move. Pictorial cues involve the information about depth that can be perceived even in a flat painting. Included are lines of perspective, the blocking of one object by another, shadows and shading and the variation in the density of textures with distance.

Convergence and retinal disparity play a role in most perceptions of three dimensions. They invoke the concept of the visual axis, or line of sight, which is an imaginary line connecting an object with its image on the retina when you look directly at the object. Suppose you look directly at an object *A*. Its image lies on the visual axis, and so on the same part of the retina in each eye, enabling the brain to fuse the two views into a single perception. The angle between the two visual axes is called the angle of convergence. It is related to the angle through which the eyes must turn in order to direct their axes at *A*. The visual system associates that angle with distance to the object: the larger the angle is, the closer the object seems to be.

When you look directly at *A*, the images of a more distant object *B* are at different places on each retina. The visual system recognizes this disparity as a cue to the depth between *A* and *B*. The recognition can also be explained in terms of convergence angles. If you look directly at *B*, the angle between the visual axes is smaller than it was for *A*. Therefore *B* must be farther away than *A*.

Retinal disparity depends partly on the separation between the two eyes,



Terry Pope's hyperscope **(left)** ***and pseudoscope*** **(right)**

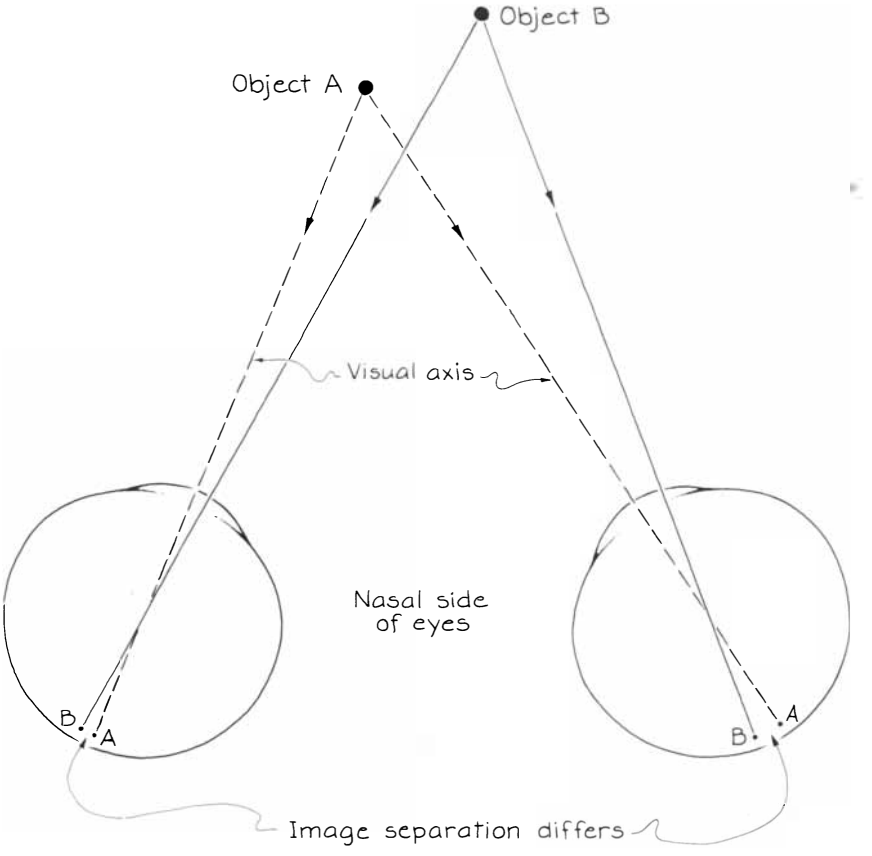
which is about six centimeters. Pope set out to explore what happens to perception if the separation is changed. One of his instruments, the hyperscope, effectively increases the separation to about 20 centimeters by means of reflections from mirrors. Similar instruments were studied by Charles Wheatstone and David Brewster in the 19th century.

The increase in the effective distance between the eyes increases the retinal disparity of images formed on the retinas and the difference in convergence angles when you look from one object to another at a different distance. Suppose you look at *A* through the hyperscope while *B* is also in view. The new disparity of separation between the images of the two objects on the retinas forces you to perceive greater depth between them. You also perceive greater depth because the difference in convergence angles for the objects is now greater.

The hyperscope also alters the apparent height and width of nearby objects. In normal vision you are accustomed to a certain relation between the size of an object's image on the retina and the object's distance, as implied by the convergence of the eyes when you look at it. Seen through the hyperscope, an object looks smaller because the angle of convergence required to see it through the mirrors is larger than normal.

Many other familiar objects take on a strange appearance through the hyperscope. For example, a person's face looks thinner and seems to have a prominent nose. All the objects immediately return to their normal appearance if you close one eye while still looking through the instrument with the other one. Because you are no longer able to compare retinal disparity or convergence angles between the eyes, you are left with only the pictorial cues about depth.

Another of Pope's instruments, the pseudoscope, makes use of mirrors to switch what the eyes are seeing. The exchange reverses the cues about distance from retinal disparity, sometimes causing a distant object to seem closer than a nearer one. The exchange of depth is most vivid for me when I look through the pseudoscope at complex arrays such as trees or brush. Branches at the rear of a tree seem closer than branches at the front. The sight is eerie because I realize that the front branches partially block my view of the rear branches. Depth is also inverted when I look at an object that can easily be reversed mentally. For example, a pot hung bottom out on the kitchen wall suddenly appears to bulge inward rather than outward.



Retinal disparity

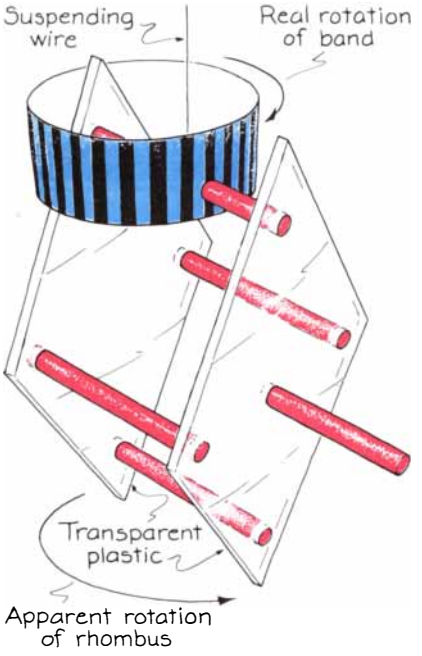
Pope has made several constructions of transparent plastic that seem to move surprisingly when a pseudoscope inverts them. One is a rhombus consisting of two plastic parallelograms held together by four metal rods. The rhombus is suspended by a thin wire. Propped on top of the rhombus is a band of alternating green and black stripes. One of the rods passes through the band to hold it in place.

I suspended this device from a ceiling lamp fixture, rotated the rhombus to twist the wire, released it and then viewed it through the pseudoscope from a distance of about five meters. Initially the rhombus and the band rotated together, first one way and then the other as the wire twisted and untwisted. Suddenly the rhombus inverted front for rear. Thereafter it and the band appeared to rotate in opposite directions. Although I knew the two objects were firmly connected, I could not shake the illusion of opposite rotation until I closed one eye.

Readers interested in buying the hyperscope, the pseudoscope or a device called the duoscope that combines the two instruments should write to Pope at Scope Productions, 102 Newbury, Berkshire, U.K. RG16 9HJ. Pope can specially design large versions suitable for museums.

Some printed advertisements invoke a sense of depth with adjoining regions of different colors. For example, small red letters on a blue background seem to me to be higher than the background if the illumination is bright. The illusion becomes stronger when I move the pattern farther away. As the illumination is made dimmer, the letters seem to drop to the plane of the background and then move below it. In very dim illumination the sense of depth disappears. Colors intermediate to the red and blue ends of the visible spectrum give rise to weaker sensations of depth.

The illusion of depth in colored patterns is due to the spreading of colors by the eye (a phenomenon called chromatic dispersion or aberration) and to the fact that the center of the pupil is not on the visual axis passing through the pupil. When rays of light pass through the curved cornea of the eye, they are refracted and their component colors are spread. The refraction is measured in relation to a line (the

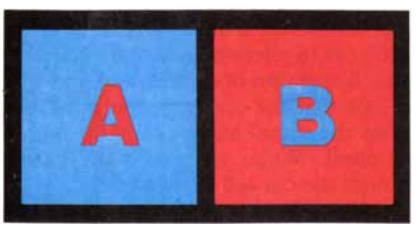


Pope's rhombus

normal) perpendicular to the surface at the point of refraction. If the incident ray is white, the refracted blue ray is closer than the refracted red ray to the normal. Intermediate colors are refracted at correspondingly intermediate angles.

Because the cornea is curved, the normals to it are oriented in different directions at different points on its surface. Hence rays entering the nasal side of the cornea are refracted in directions different from those of identical rays entering the temple side. The extent of dispersion at any location depends on the angle between the incident ray and the normal there. A larger angle yields a greater dispersion. Additional dispersion arises in the lens of the eye.

To demonstrate the illusion, which is called the color-stereoscopic phenomenon, paste two small circles, one blue and the other red, on a black card. I cut the circles from cards displaying pure tone colors as guides in color printing and photography. Inked circles from ordinary pens may serve,



The color-stereoscopic phenomenon

but impure colors weaken the illusion. Space the circles closely, putting the red one on the left. Hold the card in bright light and look directly at the red circle, making whatever adjustments are needed to bring it into focus. At each eye rays of light from the circle are refracted at the cornea, pass through the pupil and cross, forming a sharp image at the point where the visual axis intersects the retina. If the second circle were also red, it would form a sharp image slightly to one side of the first image. When that circle is blue, the image is blurry. The blue rays are additionally refracted, and so they cross in front of the retina. By the time they reach the retina they are spreading. That is the reason they produce a blurry image.

For two reasons the center of the blurry blue image does not fall on the sharp image. First, in bright illumination the pupil is small and its center is on the temple side of the visual axis. Light rays can enter the eye farther off the axis on that side than on the other side, thereby putting more of the blurry image on the nasal side of the axis. Second, additional bias in the position of the blurry image results from the skewed pathway of the visual axis through the lens of the eye. The line of symmetry through the lens lies on the temple side and at an angle of about five degrees to the visual axis. Light passing through the temple side of the lens is dispersed more than light that passes through the nasal side. The arrangement of the blue image on the nasal side of the red image creates the illusion that the blue circle is farther away than the red one.

The apparent depth is not altered if you switch your gaze and focus it sharply on the blue circle. In each eye the blue image then lies on the point where the visual axis intercepts the retina. The red rays, which are less strongly refracted, tend to focus behind the retina. When they reach the retina, they form a blurry image. The center of this image lies on the temple side of the blue image. Again the blue circle seems to be farther away because the blue image still lies on the nasal side of the red image.

The relative positions of the colored images change in decreasing illumination because each pupil then widens eccentrically and the center of the pupil shifts toward the nasal side of the eye. When the center of the pupil reaches the visual axis, the center of the blurry image almost coincides with the sharp image. The circles seem to be equidistant from the viewer.

With a further decrease in illumination each pupil widens more and its center moves off the visual axis toward the nasal side of the eye. This migration moves the blue image across the retina toward the temple side of the red one. The visual system's perception of depth is then reversed from what it was initially.

Many observers cannot discern any depth in arrangements of this kind, whereas others see the letters consistently above or below the background even when the illumination level is varied. I suspect that in the first group the observer's knowledge that the circles are equidistant overrides the illusion. In the group that does perceive a disparity the center of the pupil may stay on one side of the visual axis or the other regardless of the change in the pupil's width.

In 1965 B. N. Kishto, a resident of the Indian Ocean island of Mauritius, described in *Vision Research* ways to demonstrate chromatic dispersion by illuminating only one side of each eye. To repeat one of his demonstrations I block the nasal half of each eye's field of view with a card as I look at the colored circles. The light entering the eye on the uncovered temple side is refracted. Blue light is directed more than red toward the normal at the corneal surface. Dispersion is enhanced by the lens of the eye. The blue image ends up on the nasal side of the red image. The blue circle looks farther away than the red one.

Next I block the temple half of each eye's field of view. Although blue is still refracted more than red, the new orientation of the normal changes the relative locations of the images. The blue image is now on the temple side of the red image, and as a result the blue circle now appears to be closer than the red one.

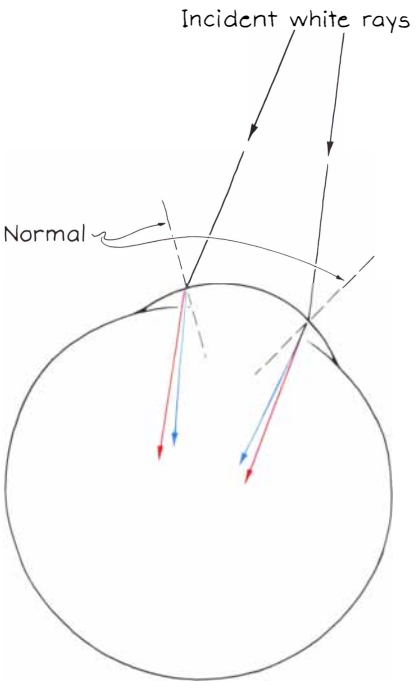
The same results can be obtained with a pinhole. Make a hole about two millimeters in diameter in an index card. Look at the colored circle with one eye closed and the pinhole in front of the other eye. The blue circle seems farther away than the red one when you position the pinhole on the temple side of the open eye. The apparent depth is reversed when the pinhole is on the nasal side.

The pinhole creates another illusion. Look through the pinhole and move it toward the temple and nasal sides of your view as far as you can without losing sight of the circles. The lateral (left-right) separation between the circles changes, being smallest when the pinhole is at its extreme nasal position and largest when it is at its extreme temple position. The increase in dispersion on the temple side results from the skewed alignment of the visual axis

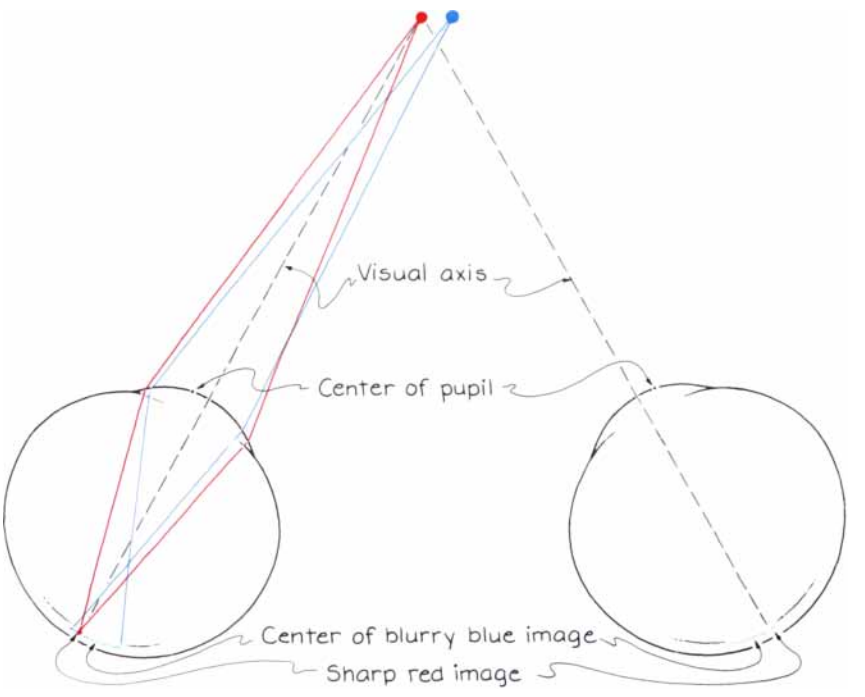
with the line of symmetry passing through the lens.

Kishto also showed how the dispersion of light can produce colored fringes on a small white patch with a dark surround when the patch is viewed through a pinhole. To test this result I pasted a small white rectangle on a black card. When I hold the pinhole on the temple side of an eye, that side of the white patch is fringed with blue and the nasal side with red. When I move the pinhole to the nasal side of the eye, the colors reverse.

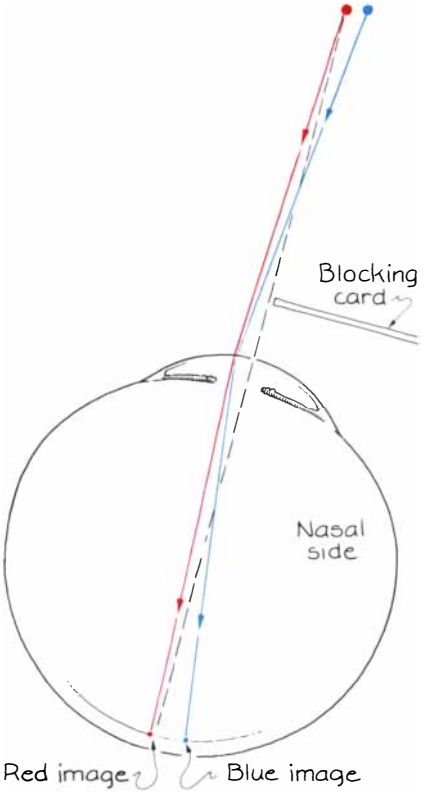
White light from the patch is dispersed into colors when it enters the eye. In the center of the retina, however, the color from one point of the patch overlaps colors from other points to re-create a perception of white. You therefore see color only

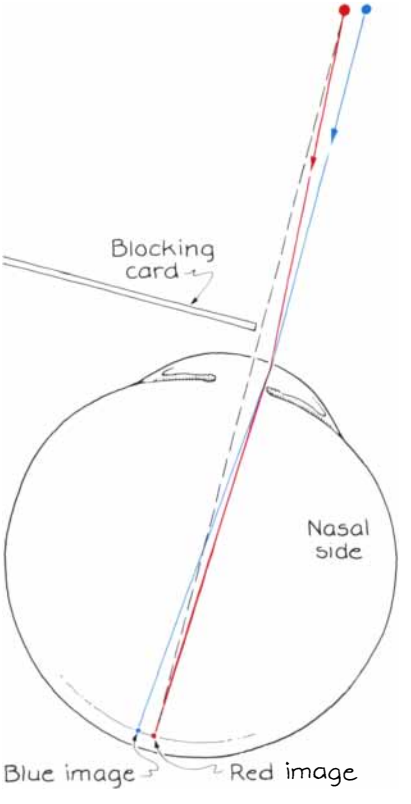


Chromatic dispersion

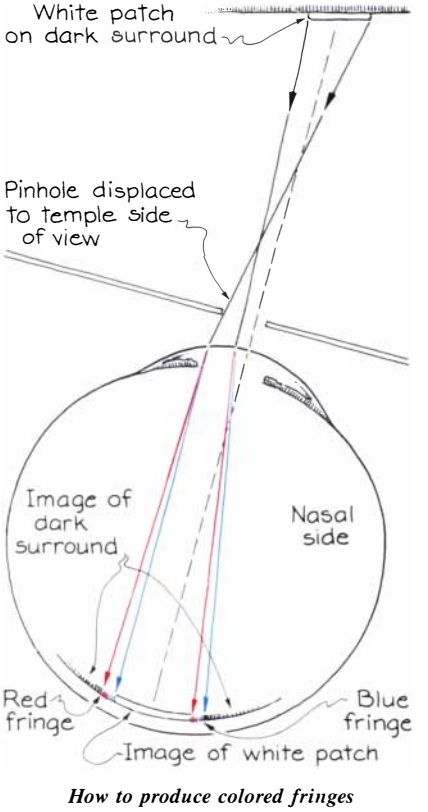


How the color-stereoscopic illusion arises

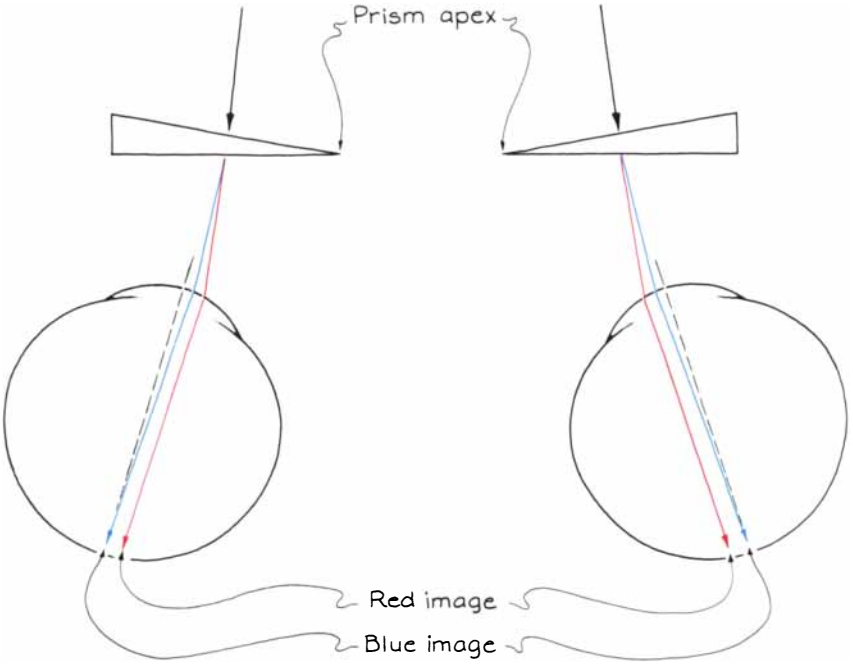




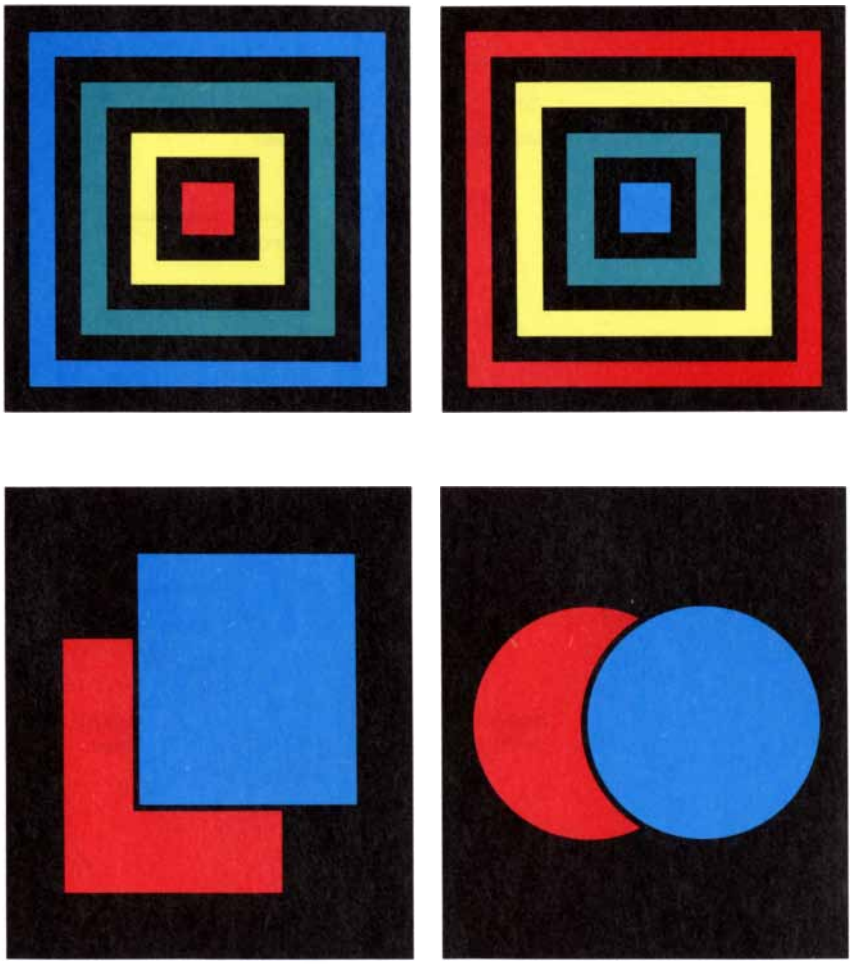
Experiments with a blocking card



How to produce colored fringes



Testing the depth illusion with prisms



Patterns designed to give pictorial cues

along the left and right edges of the rectangular patch.

In several other experiments Kishto described, observers viewed equidistant colored patches through narrow glass prisms in order to increase the dispersion of light. When the apexes of the prisms pointed toward the temple side of each eye, an observer had to decrease the convergence of his eyes to see the patches. In this situation a red patch seemed closer than a blue one. The situation was reversed when the apexes pointed toward the nasal side of each eye.

The illusion of depth worked best when the patches occupied about one degree in the field of view and were separated by about a quarter of a degree to half a degree. In some trials an observer viewed an array of four patches of different colors. With certain color arrays the inner corners of the patches seemed to curl up or down in relation to the plane of the outer corners. The illusion of depth is weaker at the more distant outer corners. Joseph Hodych of the Memorial University of Newfoundland has suggested that contour maps of magnetic fields in the ground be coded in colors and then viewed through a large magnifying lens so that the relative strength of the fields is represented by the perceived depths of the colors. The lens is in effect Kishto's first arrangement of two prisms.

In another set of experiments observers viewed arrangements of colored and black strips designed to give pictorial cues about depth. In one case the order of the colored regions from the central square outward was red, yellow, green and blue. When this pattern was viewed with both eyes looking through prisms whose apexes were directed toward each other or with one eye looking through a pinhole on the nasal side of the eye, it resembled a hallway. When the experimenter reversed the order of the colors, the pattern looked like a bellows projecting toward the observer.

If the pattern is colorless, its depth cues are ambiguous. I can see it as either a hallway or a projection toward me. With the extra sense of depth generated by the colors I become locked into one of those perceptions. With another pattern [*see bottom illustration at left*] the pictorial cue that one object is blocking my view of a more distant object is so strong that the color-stereoscopic phenomenon cannot change my perception of depth in the pattern. You might enjoy investigating other patterns where pictorial cues enhance or oppose the illusion of depth created by adjacent colors.

BREAKTHROUGH: BAKE COMPOSITES AND NEVER BOTCH A BATCH.

Strong, lightweight carbon composites that resist heat are essential for high performance jet aircraft.

Producing consistent, high-quality parts with the material, however, has been a delicate and somewhat unpredictable process. Layers of strong, light composite fabric must be slowly and skillfully baked and cured into precise aerodynamic shapes.

To make sure that every batch turns out right, we've developed a fiber-optic sensor probe to measure quality right in the curing ovens so that our technicians know exactly how fast the cure is progressing and when each batch of parts is perfectly done. Inconsistencies from batch to batch are minimized. Optimum properties are maintained. Waste is cut. Time and money are saved.

We're creating breakthroughs that make a difference in the way things work and the way people live.

We're McDonnell Douglas.

For more information, write: Advanced Materials, McDonnell Douglas, Box 14526, St. Louis, MO 63178

MCDONNELL DOUGLAS

MILITARY & COMMERCIAL AIRCRAFT SPACECRAFT & MISSILES HEALTHCARE SYSTEMS INFORMATION SYSTEMS HELICOPTERS FINANCING

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

STAR FLEET TECHNICAL MANUAL. Designed by Franz Joseph. Ballantine Books, 1975.

MICROCOMPUTER DISPLAYS, GRAPHICS, AND ANIMATION. Bruce A. Artwick. Prentice-Hall, Inc., 1985.

THE SHADOW MARKET IN HOUSING

1980 CENSUS OF HOUSING, VOL. 4: CHARACTERISTICS OF HOUSING UNITS. COMPONENTS OF INVENTORY CHANGE, PART I: UNITED STATES AND REGIONS. U.S. Department of Commerce, Bureau of the Census. U.S. Government Printing Office, 1983.

RENTAL HOUSING IN THE 1980s. Anthony Downs. Brookings Institution, 1983.

ACCESSORY APARTMENTS IN SINGLE-FAMILY HOUSING. Martin Gellen. Center for Urban Policy Research, Rutgers University, 1985.

ENGINEERING VOYAGER 2'S ENCOUNTER WITH URANUS

THE VOYAGER SPACECRAFT. R. L. Heacock in *Proceedings of the Institution of Mechanical Engineers,* Vol. 194, No. 28, pages 221–224; 1980.

VOYAGER FLIGHT ENGINEERING: PREPARING FOR URANUS. W. I. McLaughlin and D. M. Wolff. American Institute of Aeronautics and Astronautics Publication No. 85–0287; 1985.

MISSION SYSTEMS. Special issue of *Journal of the British Interplanetary Society,* Vol. 38, No. 10; October, 1985.

VOYAGER 2 MISSION TO URANUS. Special issue of *Science,* Vol. 233, No. 4759; July 4, 1986.

STUDYING THE EARTH BY VERY-LONG-BASELINE INTERFEROMETRY

THE EARTH'S VARIABLE ROTATION: GEOPHYSICAL CAUSES AND CONSEQUENCES. Kurt Lambeck. Cambridge University Press, 1980.

VERY-LONG-BASELINE RADIO INTERFEROMETRY: THE MARK III SYSTEM FOR GEODESY, ASTROMETRY, AND APERTURE SYNTHESIS. Alan E. E. Rogers et al. in *Science,* Vol. 219, No. 4580, pages 51–54; January 7, 1983.

VARIATIONS IN THE ROTATIONS OF THE EARTH. W. E. Carter, D. S. Robertson, J. E. Pettey, B. D. Tapley, B. E. Schutz, R. J. Eanes and Miao Lufeng in *Science,* Vol. 224, No. 4652, pages 957–961; June 1, 1984.

GEODETIC RADIO INTERFEROMETRIC SURVEYING: APPLICATIONS AND RESULTS. W. E. Carter, D. S. Robertson and J. R. MacKay in *Journal of Geophysical Research,* Vol. 90, No. B6, pages 4577–4587; May 10, 1985.

RNA AS AN ENZYME

SELF-SPLICING RNA: AUTOEXCISION AND AUTOCYCLIZATION OF THE RIBOSOMAL RNA INTERVENING SEQUENCE OF TETRAHYMENA. Kelly Kruger, Paula J. Grabowski, Arthur J. Zaug, Julie Sands, Daniel E. Gottschling and Thomas R. Cech in *Cell,* Vol. 31, No. 1, pages 147–157; November, 1982.

THE RNA MOIETY OF RIBONUCLEASE P IS THE CATALYTIC SUBUNIT OF THE ENZYME. Cecilia Guerrier-Takada, Kathleen Gardiner, Terry Marsh, Norman Pace and Sidney Altman in *Cell,* Vol. 35, No. 3, Part 2, pages 849–857; December, 1983.

BIOLOGICAL CATALYSIS BY RNA. T. R. Cech and B. L. Bass in *Annual Review of Biochemistry,* Vol. 55, pages 599–629; 1986.

THE INTERVENING SEQUENCE RNA OF *TETRAHYMENA* IS AN ENZYME. Arthur J. Zaug and Thomas R. Cech in *Science,* Vol. 231, No. 4737, pages 470–475; January 31, 1986.

THE HIGGS BOSON

GAUGE THEORIES OF THE FORCES BETWEEN ELEMENTARY PARTICLES. Gerard 't Hooft in *Scientific American,* Vol. 242, No. 6, pages 104–138; June, 1980.

CONSTRUCTING QUARKS: A SOCIOLOGICAL HISTORY OF PARTICLE PHYSICS. Andrew Pickering. University of Chicago Press, 1984.

THE SECOND CREATION: MAKERS OF THE REVOLUTION IN TWENTIETH CENTURY PHYSICS. Robert P. Crease and Charles C. Mann. Macmillan Publishing Company, 1985.

ANTARCTIC FISHES

COASTAL AND DEEP-WATER BENTHIC FISHES OF THE ANTARCTIC. Hugh H. DeWitt in *Antarctic Map Folio Series, Folio 15,* edited by Victor C. Bushnell. American Geographical Society, New York, 1971.

BIOLOGICAL ANTIFREEZE AGENTS IN COLDWATER FISHES. Arthur L. DeVries in *Comparative Biochemistry and Physiology,* Vol. 73A, No. 4, pages 627–640; 1982.

THE EVOLUTION OF NEUTRALLY BUOYANT ANTARCTIC FISHES: THEIR SPECIALIZATIONS AND POTENTIAL INTERACTIONS IN THE ANTARCTIC MARINE FOOD WEB. J. T. Eastman in *Antarctic Nutrient Cycles and Food Webs (Proceedings of the Fourth SCAR Symposium on Antarctic Biology),* edited by R. W. Siegfried, P. R. Condy and R. M. Laws. Springer-Verlag, 1985.

FEATURES AND OBJECTS IN VISUAL PROCESSING

SEEING: ILLUSION, BRAIN, AND MIND. John P. Frisby. Oxford University Press, 1980.

FORM AND ORGANIZATION. Irwin Rock in *Perception.* Scientific American Books, Inc., 1984.

PROPERTIES, PARTS AND OBJECTS. A. Treisman in *Handbook of Perception and Performance: Vol. 2,* edited by K. Boff, L. Kaufman and J. Thomas. John Wiley & Sons, Inc., 1986.

GALILEO AND THE SPECTER OF BRUNO

GIORDANO BRUNO AND THE HERMETIC TRADITION. Frances A. Yates. University of Chicago Press, 1964.

GALILEO GALILEI: A BIOGRAPHY AND INQUIRY INTO HIS PHILOSOPHY OF SCIENCE. Ludovico Geymonat. McGraw-Hill Book Company, 1965.

GALILEO AND THE LONG SHADOW OF BRUNO. Edward A. Gosselin and Lawrence S. Lerner in *Archives Internationales d'Histoire des Sciences,* Vol. 25, No. 97, pages 223–246; December, 1975.

THE ASH WEDNESDAY SUPPER (LA CENA DE LE CENERI). Edited and translated by Edward A. Gosselin and Lawrence S. Lerner. Archon Books, 1977.

GALILEO. Stillman Drake. Oxford University Press, 1980.

THE AMATEUR SCIENTIST

THE COLOUR STEREOSCOPIC EFFECT. B. N. Kishto in *Vision Research,* Vol. 5, pages 313–329; June, 1965.

THE EFFECT OF PUPIL SIZE VARIATIONS ON THE COLOUR STEREOSCOPIC PHENOMENON. Jon Martin Sundet in *Vision Research,* Vol. 12, pages 1027–1032; May, 1972.

BREWSTER AND WHEATSTONE ON VISION. Edited by Nicholas J. Wade. Academic Press, 1983.

How we pushed America's premier road sedan to the ultimate braking point.

When driving conditions are less than ideal, Pontiac 6000 STE helps keep you out of a bind with advanced anti-lock braking. This new system delivers improved braking capability on almost any surface, wet or dry. Together with superb handling and fuel-injected V6 power, it gives you a machine with some of the finest performance credentials going. Or stopping.

PONTIAC 6000

WE BUILD EXCITEMENT