SCIENTIFIC AMERICAN

NOVEMBER 1989
$2.95

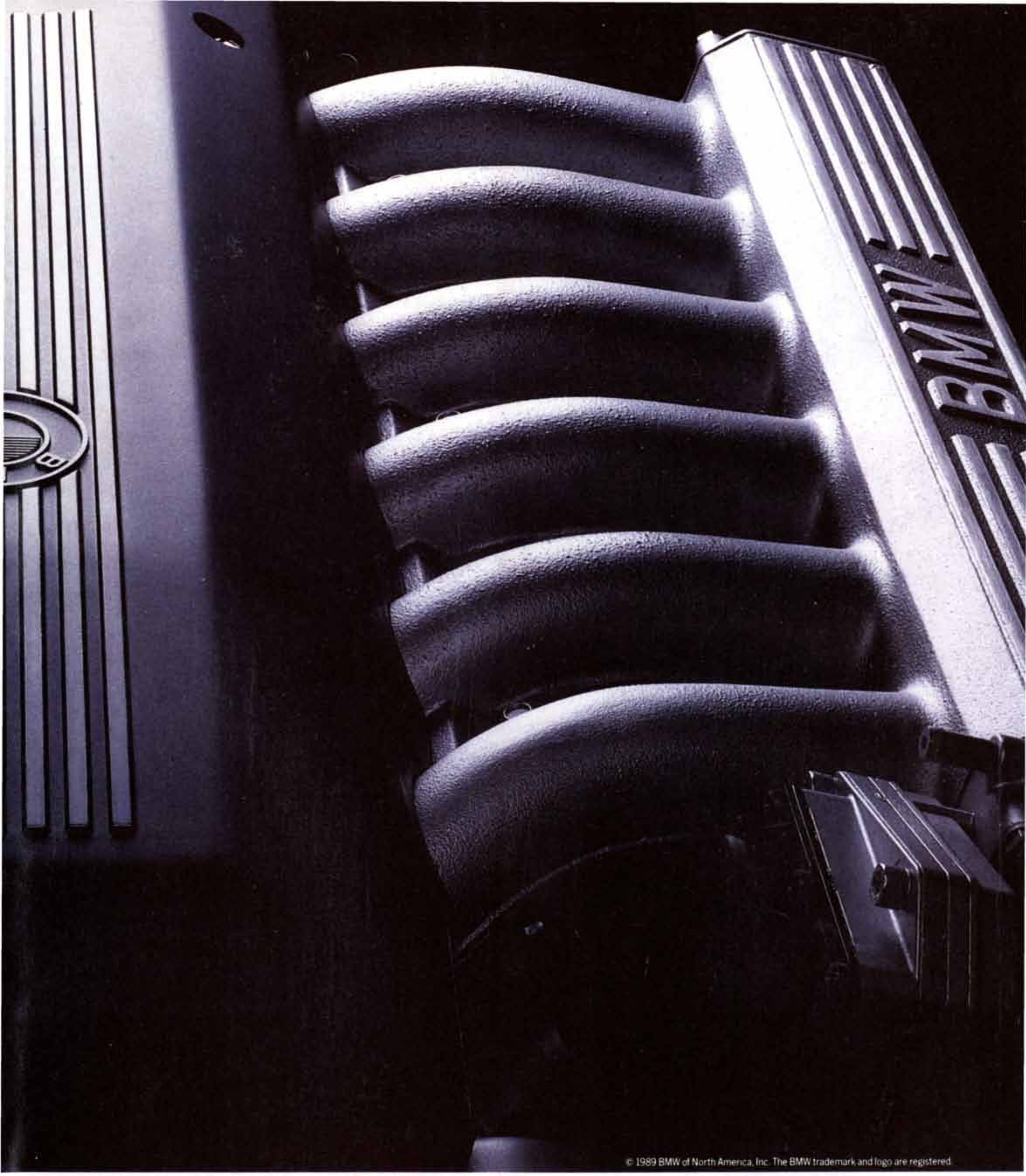
Neptune portfolio: spectacular visit to a turbulent planet.

Double-beta decay: observing the undetectable.

Yellowstone: the fires are out, the debate burns on.

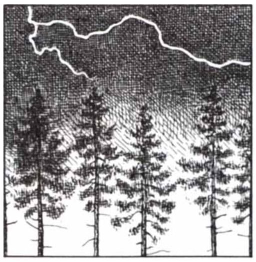
Shuttle glow: *250 kilometers above the earth an attenuated wind raises a halo of light from the spacecraft's skin.*

It will tell you whether they were limited by cost constraints and timetables, or only by their imaginations. It will tell you whether they are simply content with compromise, or resolute about perfection — people who are passionate about spending months just to perfect a single component. Looking carefully at the engine of your car is the


BMW
© 1989 BMW of North America, Inc. The BMW trademark and logo are registered.

SCIENTIFIC AMERICAN

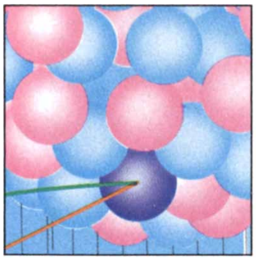
November 1989 Volume 261 Number 5

36 

The Yellowstone Fires

William H. Romme and Don G. Despain

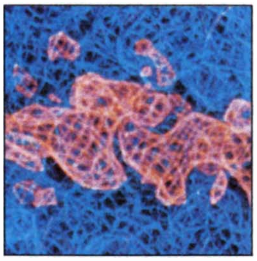
During the summer of 1988 flames swept across much of Yellowstone National Park. More acres burned that year—720,000 to be exact—than had burned in any decade since 1872. Could the fires have been prevented? Should they have been? No to both questions. Large fires, it seems, are not only inevitable; they are also necessary mediators of ecological change.

48 

Double-Beta Decay

Michael K. Moe and Simon Peter Rosen

Experiments confirm what has been predicted by theory: that two neutrons will decay simultaneously into two protons, two electrons and two neutrinos. Physicists hope that by studying the rare phenomenon—known as double-beta decay—they may discover whether or not the neutrino has mass; if it does, the implications for cosmic evolution could be profound.

56 

How T Cells See Antigen

Howard M. Grey, Alessandro Sette and Søren Buus

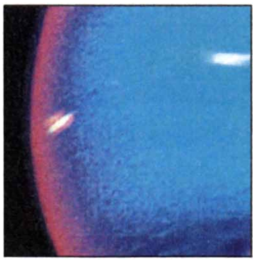
These small white blood cells marshall the body's immune response and destroy infected cells. Yet by themselves they are blind to signs of invasion. Other cells, it turns out, must break down foreign material and "present" it to the *T* cells. Understanding this intricate process may lead to new vaccines and treatments for immune disorders.

68 

The Mammalian Choroid Plexus

Reynold Spector and Conrad E. Johanson

By selectively regulating what passes out of the bloodstream and into the cerebrospinal fluid, the choroid plexus acts as a gatekeeper to the brain. New findings underscore the importance of this small cluster of cells. They also open up possibilities for a new generation of drugs that can cross the blood-brain barrier and attack pathogens in the brain.

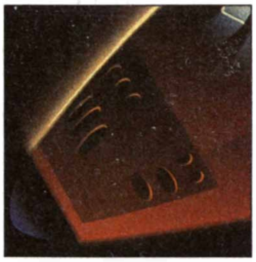
82 

SCIENCE IN PICTURES

Neptune

June Kinoshita

Twelve years after being launched, the intrepid spacecraft *Voyager 2* has reached Neptune, the final destination of its planetary odyssey. The reward is a stunning portfolio of images showing a surprisingly dynamic planet.

92 

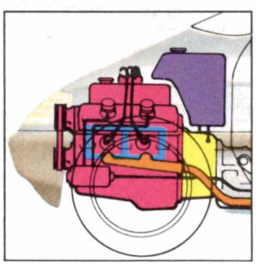
Shuttle Glow
Donald E. Hunton

During flight a strange orange glow emanates from the tail and engine pods of the U.S. space shuttle. The glow itself may be benign but is associated with something more sinister. Atmospheric gases, it seems, react chemically with the shuttle's outer skin, causing it to erode. If unchecked, surface erosion can threaten the structural integrity of a spacecraft.

100 

The World's Oldest Road
John M. Coles

For 6,000 years, a wood walkway lay buried in a peat bog in southwestern England, protected and preserved by the acid and water of the bog. The excavated wood, together with artifacts and plant remains recovered from the surrounding peat, yields a remarkably detailed picture of life in an early Neolithic community.

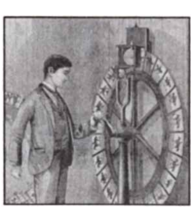
108 

The Case for Methanol
Charles L. Gray, Jr., and Jeffrey A. Alson

Spurred by concerns of declining air quality, the race for alternative automotive fuels is on. The authors advocate a switch to methanol, arguing that it is the most economic and environmentally prudent of the fuels now being considered. How should the U.S. respond? By developing a methanol-based transportation system—the authors say—as soon as possible.

DEPARTMENTS

8 Letters

14 

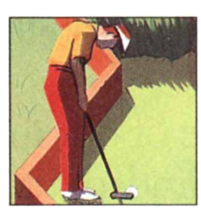
50 and 100 Years Ago

1889: An iron wheel whirls photographs to create a moving image.

17 Science and the Citizen

76 Science and Business

116 The Amateur Scientist

120 

Computer Recreations

The *Scientific American* Open: devising a game of microgolf.

124 Books

130 Essay: *Daniel Patrick Moynihan*

Towers of babble.

What we have here, communicators, is a failure to communicate.

An electronic cacophony of disparate subnetworks—PBX's, LAN's, T1's. All working. Each with its own language and agenda.

To handle this information management nightmare, scientists at our NYNEX Science & Technology Center are developing the software system of the future.

Manager of Managers.

From one powerful workstation, the system provides global management, using artificial intelligence to isolate failures and reduce subnetwork downtime.

This is only one of our insights into emerging information technologies. For now-minded futurists, a state-of-the-art NYNEX network management solution can answer needs like call accounting, bill verification and configuration management, as well as network planning and design.

THE COVER painting is a head-on view of the space shuttle as it flies through the earth's highly reactive uppermost atmosphere (see "Shuttle Glow," by Donald E. Hunton, page 92). The impact of oxygen atoms and nitrogen molecules on the shuttle's surface leads to the formation of excited compounds that emit a diffuse yellow-red glow. Areas without the glow consist of materials that react chemically with the oxygen atoms and are therefore being eroded.

THE ILLUSTRATIONS

Cover painting by George V. Kelvin

Page	Source
36	Gary Braasch
38-41	Tom Prentiss
42	Johnny Johnson
43	William H. Romme
44	Gary Braasch
49-54B	Gabor Kiss
57	Morten H. Nielsen and Ole Werdelin, University of Copenhagen
58-60	George V. Kelvin
61	Mark M. Davis and Pamela J. Bjorkman, Stanford University
62-63	George V. Kelvin
64	Mark A. Saper, Harvard University (*top*), George V. Kelvin (*bottom*)
69-73	Carol Donner
74	Johnny Johnson
82-83	National Aeronautics and Space Administration/Jet Propulsion Laboratory, California Institute of Technology
84	Hank Iken (*top*), NASA/Jet Propulsion Laboratory (*bottom*)
85-90	NASA/Jet Propulsion Laboratory
93-94	George V. Kelvin
95	James T. Visentine, National Aeronautics and Space Administration's Lyndon B. Johnson Space Center
96	George V. Kelvin
98	Gary Swenson, Lockheed Palo Alto Research Laboratories
101	Patricia J. Wynne (*left*), John M. Coles (*right*)
102-104	Patricia J. Wynne
105	John M. Coles
106	Patricia J. Wynne
109-111	George Retseck
112-113	Johnny Johnson
114	George Retseck
116-119	Michael Goodman
120	Russell Ruthen
121	Edward Bell
122	Andrew Christie

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 rates: one year $27, two years $48, three years $66 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51593.

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Armand Schwab, Jr., *Managing Editor;* Timothy Appenzeller, Laurie Burnham, *Associate Editors;* Timothy M. Beardsley; Elizabeth Corcoran; John Horgan; June Kinoshita; Philip Morrison, *Book Editor;* John Rennie; Ricki L. Rusting; Russell Ruthen; Paul Wallich; Karen Wright

ART: Samuel L. Howard, *Art Director;* Murray Greenfield, *Associate Art Director;* Edward Bell, *Assistant Art Director;* Johnny Johnson

COPY: Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Michele S. Moise

PRODUCTION: Richard Sasso, *Vice-President Production and Distribution; Managers:* Carol Eisler, *Manufacturing and Distribution;* Carol Hansen, *Electronic Composition;* Leo J. Petruzzi, *Manufacturing and Makeup;* Carol Albert; Madelyn Keyes; William Sherman

CIRCULATION: Bob Bruno, *Circulation Director;* Lorraine Terlecki, *Business Manager*

ADVERTISING OFFICES: NEW YORK: Scientific American, 415 Madison Avenue, New York, NY 10017; Robert F. Gregory, *Advertising Manager;* Lisa Carden; Peter Fisch; John Grant; Meryle Lowenthal; William Lieberman, Inc. CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager;* Litt Clark, *Midwest Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; William F. Moore, *Advertising Manager;* Edward A. Bartley, *Detroit Manager.* WEST COAST: 1650 Veteran Avenue, Suite 101, Los Angeles, CA 90024; Kate Dobson, *Advertising Manager;* Joan Berend, San Francisco. ATLANTA, BOCA RATON: Quenzer/Stites. CANADA: Fenn Company, Inc. DALLAS: Griffith Group.

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

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; GWP, Düsseldorf. HONG KONG/SOUTHEAST ASIA: C. Cheney & Associates. SEOUL: Biscom, Inc. SINGAPORE: Cheney Tan Associates. TOKYO: Nikkei International, Ltd.

PUBLISHER: John J. Moeling, Jr.

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

PRESIDENT AND CHIEF EXECUTIVE OFFICER: Claus-Gerhard Firchow

EXECUTIVE COMMITTEE: Claus-G. Firchow; *Executive Vice-President and Chief Financial Officer,* R. Vincent Barger; *Vice-Presidents:* Linda Chaput, Jonathan Piel, Carol Snow

CHAIRMAN OF THE BOARD: Georg-Dieter von Holtzbrinck

CHAIRMAN EMERITUS: Gerard Piel

"Sure, Canon's the most popular copier in America. But I bet you didn't know Canon has a personal copier that's perfectly suited to family life.

It's got a price tag even a mother — or father — could love. Really affordable. And it's simple enough for even a little kid to use.

And talk about reliable! It's just about maintenance-free. And definitely worry-free! Count on beautiful copies, copy after copy after copy. *That's* all because of Canon's special pop-in, pop-out, single cartridge system.

You know it'll save you running to the copy shop with tax info and insurance bills. But there's a ton of new things you'll do with it, too. Your own special greeting cards. Kids' school projects. Stuff like that.

It's so nice to have a copier around the house. And now, thanks to Canon, at last it's possible."

Canon PC PERSONAL COPIERS

Canon U.S.A., Inc., Home Office Products Division. One Canon Plaza, Lake Success, N.Y. 11042. Regional phone numbers: New York (516) 488-6700, Chicago (312) 250-6200, Atlanta (404) 448-1430, Dallas (214) 830-9600, Los Angeles (714) 979-6000. Enjoy easy extended payments with the Canon Credit Card. Ask for details at participating Canon dealers and retailers. Available only in U.S. © 1989 Canon U.S.A., Inc.

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 G. De Alessandri, 11
20144 Milano, Italy
Advertising correspondence:
Publietas, S.p.A., Via Cino de Duca, 5,
20122 Milano, Italy

サイエンス
Y950/copy Y10,440/year Y14,000/(abroad)
Editorial, subscription, advertising correspondence:
Nikkei Science, Inc.
No. 9-5, 1-Chome, Otemachi
Chiyoda-ku, Tokyo, Japan

INVESTIGACION Y CIENCIA
500 Ptas/copy 5500 Ptas/year 6200 Ptas (abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica S.A.,
Calabria, 235-239
08029 Barcelona, Spain

POUR LA SCIENCE
27FF/copy 265FF/year 315FF/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
3.80RMB/copy 45.60RMB/year $48/(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

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) 754-0550 Telex: 236115

LETTERS

To the Editors:

It comes as a surprise that there is still talk about who first conceived the geodesic dome made famous by R. Buckminster Fuller.

According to Tony Rothman, who is cited in "Surpassing the Buck" ["Science and Business," SCIENTIFIC AMERICAN, September], it was the Zeiss Optical Company that in the 1920's was the first to patent and construct a geodesic dome—well before 1954, when Fuller was granted a patent for his invention.

Yet way back in the middle of the 19th century the French architect Eugène-Emmanuel Viollet-le-Duc had already articulated the idea fundamental to the geodesic dome. Viollet-le-Duc was intrigued by the potential uses of iron in architecture. Like some of his contemporaries, he advocated the use of iron frameworks—the "skeleton" construction from which the modern skyscraper evolved.

In a series of "Lectures on Architecture" that he gave before 1860, Viollet-le-Duc spoke of the potential of iron for constructing large vaulted spaces unattainable by ancient and medieval builders. The vault he conceived had an iron framework made up of separate parts: "If we consider iron as easy to employ and connect in straight pieces," he said (Lecture XII, "On the Construction of Buildings"), "and if of these separate pieces we form a kind of independent network, and on this network of girders we rest the vaulting in separate parts, we shall thus have contrived a system of iron framework... and a method of covering wide spaces." He illustrated his idea in a sketch: "Let figure 16 represent a polyhedron capable of being inscribed within a hemisphere, and consisting of regular sides forming octagons, hexagons and squares. It is evident that if we set up a framework of iron... we shall be able to cover the various parts of this network with portions of vaulting."

Viollet-le-Duc's concepts were well known. Benjamin Bucknall translated his "Lectures" into English in 1889.

MILTON KIRCHMAN

New York, N.Y.

To the Editors:

The Pentagon sources quoted in "Ivan's Eyes" ["Science and the Citizen," SCIENTIFIC AMERICAN, June] are being economical with the truth. The claim that "the Soviet Union... started it [antisatellite weapons]" ignores the well-known early history of the U.S. ASAT program. Curious readers will find this history in chapter six of *The Militarization of Space: U.S. Policy, 1945-1984*, by Paul B. Stares (Cornell University Press, 1985). Stares describes a test carried out on February 14, 1964, in which a Thor missile successfully intercepted a Transit 2A rocket body. The next two tests in the series of four were also successful. The system became operational in June, 1964, four years before the first Soviet ASAT interception test of October, 1968. Although never tested with live warheads (it would have carried a 1.5-megaton nuclear device), the system remained available, and periodic test launches took place until 1970, when it was effectively phased out.

NICK WATKINS

Space and Plasma Physics Group
University of Sussex
England

To the Editors:

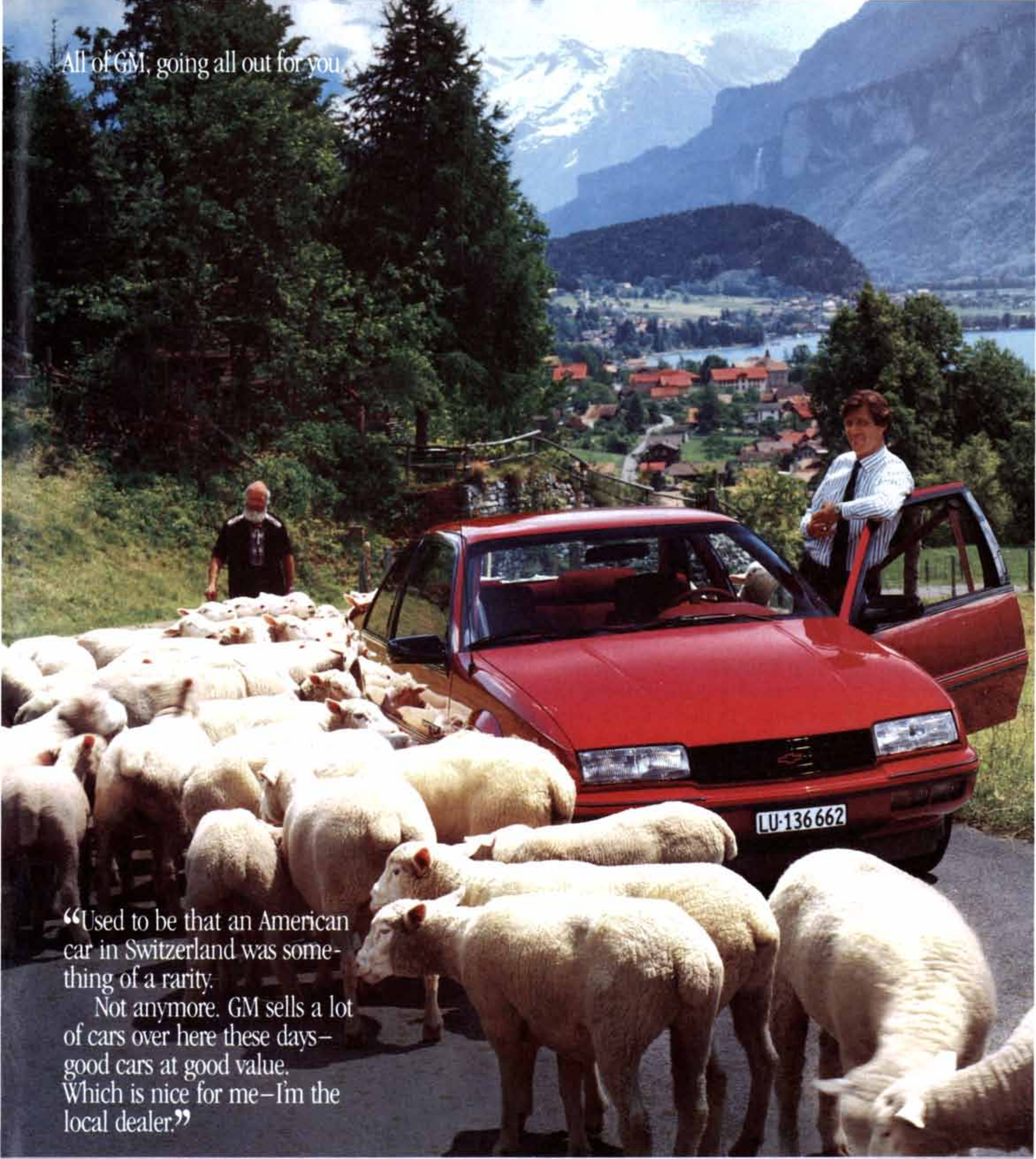
In the review of G. Venkataraman's book *Journey into Light: Life and Science of C. V. Raman* ["Books," SCIENTIFIC AMERICAN, August], you refer to C. V. Raman as "the first Asian Nobel laureate." In fact, the first Asian Nobel laureate (and the first Indian laureate) was the great Indian poet and playwright Rabindranath Tagore, who won the Nobel prize for literature in 1913—17 years before Raman won the prize in physics.

RAKESH MISRA

Chandler, Ariz.

To the Editors:

The excellence of static and dynamic design of most surviving antique structures, including the one discussed in "The Roman Aqueduct of Nîmes," by George F. W. Hauck [SCIENTIFIC AMERICAN, March], is beyond any doubt. If we stress the word *surviving*, however, the question shifts from skill to statistics. Clearly, structures that have endured two or more millennia were well planned and well executed. But considering that perhaps 1 percent or less of the structures dating from 2,000 or more years ago have survived (even if we count multiply restored ruins), should one conclude



Fact: GM exports more cars around the world from North America than any other car manufacturer.

GM
MARK OF EXCELLENCE

Chevrolet Beretta

Chevrolet, Pontiac, Oldsmobile, Buick, Cadillac, GMC Truck that 99 or more percent of the ancient architects were failures? Probably not: during 2,000 years of survival, luck and "natural selection" must also have played a role. To be sure, no poorly conceived structure could have lasted even 10 or 20 years. (Think of the first version of the choir vault of the Cathedral of Beauvais collapsing as it was built because the strength of the stone could not bear the pressures imposed on it.) But surely many well-conceived structures fell in ruins (as most modern "well engineered" edifices will) in a few tens or hundreds of years. Millennia-old structures certainly attest to the skill of their builders—but even more, they are witnesses to their own good fortune.

MICHEL BENARIE

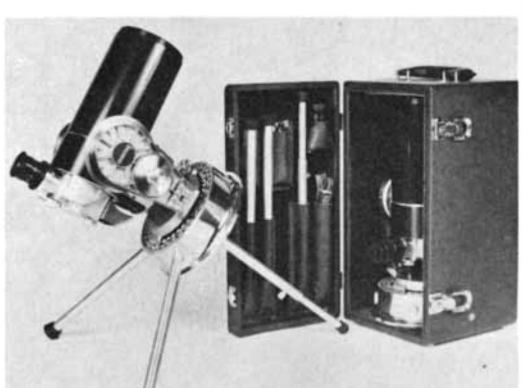
Executive Editor,
European Cultural
Heritage Newsletter
Bretigny, France

Author's response:

To be sure, the Pont du Gard's very age indicates its merits on the principle of the survival of the fittest. More than durability is involved, however, and more than statistics. I do not agree that "structures that have endured two or more millennia were [necessarily] well planned and well executed." A crude pile of rocks would last a long time, too. In contrast, Caesar Augustus built a bridge across the Rhine that lasted only 18 days before he had it destroyed, and for all I know, it might have collapsed on its own not much later; evidently, though, it was well engineered for its purpose. Structures such as the Pont du Gard, the Pantheon, Trajan's timber bridge over the Danube or the Colosseum with its canvas roof impress the engineer as no Egyptian pyramid can. Surviving or not, they fulfilled their functions economically, efficiently and elegantly (to borrow engineer David P. Billington's criteria) for as long as needed.

The Nîmes aqueduct was surely intended to be a durable investment. It did serve for 800 years, and if it had been maintained it could have lasted longer. Was the reason blind luck? I asked the question in my article, but only rhetorically. Benarie suggests that the Pont du Gard is a witness mostly to the good fortune of its builders. I rather think that we are the ones who are lucky that the great bridge was so skillfully designed.

GEORGE F. W. HAUCK



Actual Size.

63 positive-action keys with audible click and familiar QWERTY keyboard

80C88 static CMOS processor runs at 4.9152 MHz

32-pin connector for RAM/ROM Memory Cards (replace floppy disks for storing data and programs)

Measures 7.8″×4.1″×1.2.″ Weighs 15.87 oz.

Includes 5 applications: spreadsheet (Lotus® 1-2-3® file-compatible), word processing (ASCII file-compatible), telephone/address book with tone dialing, appointment diary with multiple alarms, and calculator

128K continuously powered internal RAM, externally expandable, and 256K internal ROM with DOS, BIOS, and applications

Super-twist LCD screen

256×64 pixel graphics mode; 40×8 text mode (with a moveable window on 80×25 character display)

60-pin expansion bus

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

NOVEMBER, 1939: "At first thought it would seem that the peculiar techniques of radio broadcasting—spot news, high-geared commentators, and so on—would give the whole world a grandstand seat in the event of war. But when the situation grew tense in central Europe, when fateful September 3 rolled around, the humming of the air waves brought thinking people to the realization that a communicating system was more than falling down on a job; it was providing a medium whereby misinformation strode side by side with information."

"AIRMAIL FOR SMALL TOWNS: Flying the air-ways in specially designed and equipped ships, 'sky mail clerks' are now able to make deliveries and to pick up mail without the necessity of landing. The ground equipment consists merely of two steel poles, 30 feet high, set 60 feet apart. Each pole is topped by a brilliant orange marker. Stretched between the poles, and attached by spring clips, is a transfer rope from which is suspended the mail bag to be picked up. With a relatively simple system of grapple hooks, cables and winches, it is possible to deliver and pick up mail while flying at between 90 and 110 miles an hour."

"Cancer, at least of certain types, is caused by filterable viruses. Fluids from malignant transplantable tumors of chickens were whirled in the ultracentrifuge, passed through fine-pored filters, and otherwise treated after the manner of virus-containing fluids in known animal and plant diseases. Materials obtained from these cancer-fluid filtrates, injected into the tissues of healthy chickens, produced typical cancerous growths."

"Diamonds actually fall from the sky. That real diamonds exist in meteorites has just been demonstrated for the first time by X-ray examination of hard, black crystals in a specimen from Meteor Crater in Arizona."

SCIENTIFIC AMERICAN

NOVEMBER, 1889: "Astronomers of late have been making some exceedingly interesting discoveries about the huge planet Jupiter, which is now visible in the evening sky. It is evident that a tremendous current is continually sweeping right around the huge globe of Jupiter over its equatorial regions. The cloud belts are all in motion, but not all with the same velocity. The equatorial part of the atmosphere appears to be flowing past the regions on either side of it at the rate of some 240 miles an hour! Another very singular phenomenon is the great red spot which has been seen in the planet's southern hemisphere ever since the summer of 1878. Incredible as the thought seems, there are certain facts which suggest the possibility that this phenomenon may be an elevated region of the planet thrust up through its environment of clouds."

"Edward Bellamy, the author of *Looking Backward,* has just written a fanciful sketch entitled 'With the Eyes Shut,' in which he describes an approaching phonographic age. The uses which are found for the phonograph are novel and amusing. Passengers on railway trains are supplied with phonographic literature so that their eyes are not injured by reading in a jolting coach. The names of the stations are announced by phonograph in clear tones which form a striking contrast to the incomprehensible gibberish of the ordinary brakeman."

"Eighteen tons of steel disappear every day on the single system of the London and Northwestern Railway through wear and rust."

"Mr. Ottamar Anschuetz of Lissa, Prussia, has been very successful in taking instantaneous photographs of flying birds, running horses, jumping men, etc. He has also invented an apparatus by means of which these pictures may be exhibited in a very perfect manner. This instrument (which is shown in our large engraving) is known as the electrical tachyscope. It consists of an iron wheel of sufficient diameter to hold an entire series of these pictures on its periphery. As the wheel rotates and each picture comes before a Geissler tube, it is illuminated by an electric discharge through the tube. This discharge, being instantaneous, shows each picture in an apparently fixed position, and these pictures succeed each other so rapidly that the retinal image of one picture is retained until the next is superimposed upon it, thereby giving to the observer the sense of a continuous image in constant motion."

Anschuetz's electrical tachyscope, for showing moving pictures


Bombay® Sapphire™ Gin. 47% alc/vol (94 Proof).
100% grain neutral spirits. © 1989 Carillon Importers, Ltd., Teaneck, N.J.
BOMBAY SAPPHIRE
Distilled
LONDON
DRY GIN
PRODUCT OF ENGLAND
BOMBAY SAPPHIRE.
POUR SOMETHING PRICELESS.

SCIENCE AND THE CITIZEN

20 BIOLOGICAL SCIENCES
24 PHYSICAL SCIENCES
28 MEDICINE
30 PROFILE

Not So Hot

New studies question estimates of global warming

When editors and newscasters routinely bandy about the term "global warming" without bothering to explain it, then the idea could be said to have entered the body of public knowledge, accepted by most as immutable fact.

Yet the public's acceptance of global warming caused by the greenhouse effect belies the fluid nature of the science. The conclusion that the buildup of carbon dioxide and other greenhouse gases might lead to catastrophic warming of the earth's climate continues to generate debate among scientists. As investigators refine the computer models on which global climate projections are based, some estimates of probable warming are being lowered. Moreover, new data suggest that human activity may influence the climate in ways that have previously been neglected.

Few climatologists doubt that the rapid increase of trace gases in the atmosphere resulting from human activity will produce some warming, with potentially severe consequences. Estimates of how much and how soon are made with climate models run on supercomputers. Earlier models projected that global temperature would increase from three to 5.5 degrees Celsius during the next century, but they included only crude estimates of the roles cloud cover and oceans play in mediating climatic change.

Clouds seem to have a net cooling effect, so if cloud cover increased it might, other things being equal, partly offset any warming brought on by the buildup of carbon dioxide and other greenhouse gases. Oceans, for their part, act as a vast heat sink, dissipating and circulating heat.

Two recent climate-model studies have incorporated more sophisticated representations of clouds and oceans than previous models. One such model, described in *Nature* by John F. B. Mitchell and his colleagues at the British Meteorological Office, accounts for the fact that ice clouds have different properties than water clouds do. According to the model, even if levels of carbon dioxide and other gases effectively double by the middle of the next century, the atmosphere will warm by only 1.9 degrees C, as opposed to the 5.2 degrees C predicted by a comparable model that does not distinguish between ice and water clouds.

The second model, described in *Climate Dynamics* by Warren M. Washington and Gerald A. Meehl of the National Center for Atmospheric Research in Boulder, Colo., includes more sophisticated calculations of ocean circulation and heat exchange. Their findings suggest that 30 years after an instantaneous doubling of carbon dioxide in the atmosphere (an admittedly unrealistic scenario) the earth would have warmed by only 1.6 degrees C, substantially less than models with a simpler ocean would predict.

Human activity may also affect climatic change in a way that has been previously unrecognized. Analyses of satellite data suggest that sulfate aerosols formed from pollutants (such as the sulfur dioxide that rises from urban centers) can significantly increase the earth's albedo (the light and heat that the earth and its atmosphere reflect back into space). Robert D. Cess of the State University of New York at Stony Brook says his analysis of satellite data indicates a significant increase in the albedo of stratocumulus clouds over the ocean for thousands of miles downwind of major population centers such as Southeast Asia and the east coast of North America. It is thought that the sulfate particles provide nuclei around which cloud droplets form and so increase cloud reflectance.

Similarly, Philip A. Durkee of the Naval Postgraduate School in Monterey, Calif., has found threefold increases in the reflectance of stratocumulus clouds over the ocean hundreds of miles downwind from San Francisco and Los Angeles. Such results suggest that anthropogenic increases in albedo caused by greater numbers of droplets might be large enough to affect the global radiation balance. Thomas M. L. Wigley of the University of East Anglia in England has even suggested that an increase in albedo in the Northern Hemisphere might explain why the warming trend in this century has been smaller than was predicted by earlier and more simplistic climate models.

Patrick J. Michaels of the University of Virginia, a vocal critic of what he sees as premature policy responses to global warming, points out that studies by Thomas R. Karl of the National Climatic Data Center in Asheville, N.C., show that in the U.S. and Canada during the period from 1941 to 1980 there was a marked decrease in the daily temperature range, attributable mainly to an increase in nighttime and early-morning temperatures. Michaels thinks the finding supports the notion that global cloud cover may be increasing. An increase in cloudiness would increase nighttime temperatures by trapping heat emitted by the earth but decrease temperatures during the day by increasing cloud reflectance. According to Michaels, the greenhouse effect might therefore be more benign than has been supposed—or perhaps even beneficial. Growing seasons might be lengthened and the frequency of droughts reduced. Michaels's views are not widely shared: Kevin E. Trenberth of the National Center for Atmospheric Research counters that the decreasing daily temperature range might be a consequence of urban development near measuring stations.

Another participant in the debate is the Marshall Institute in Washington, D.C. Founded in 1984 and privately funded, the institute's most notable previous contribution to public policy discussion was a stout defense of the Strategic Defense Initiative. Now under the Marshall banner, Frederick Seitz, a past president of the National Academy of Sciences, Robert Jastrow, the founder of the Goddard Institute for Space Studies and William A. Nierenberg, director emeritus of the Scripps Institution of Oceanography, have seized on another weak link in current climate models. They suggest (with no firm evidence) that fluctuations in the sun's brightness during the past 100 years may be responsible for the uneven course of global warming in this century. If so, they argue, the planet's response to the buildup of greenhouse gases has been smaller than climate models suggest, and expectations of warming should be scaled back accordingly.

Michaels thinks projections of global warming will become even more moderate as models become increasingly sophisticated and incorporate more climate feedbacks. Cess, for one, disagrees: he believes improvements to models are as likely to make projections worse as they are to make them better. Only more research is likely to settle the arguments. *—Tim Beardsley*

Ningún Nombre

Identifying Argentina's desaparecidos

In late September Argentina's new president, Carlos Saúl Menem, bowed to pressure to pardon some 20 police and military officers accused of killing dissidents during the 1970's. But some of the victims—the *desaparecidos,* or "disappeared ones"—are emerging from anonymity to bear silent witness against their executioners, if not in court then in history. The Argentine Forensic Anthropology Team, a group of seven investigators working out of a small office in Buenos Aires, is quietly extending its efforts to identify the remains of persons murdered during the military regime. In collaboration with colleagues in the U.S., the Argentine workers are developing ways to apply DNA-fingerprinting techniques to the identification of badly decomposed corpses. The team is also exporting its expertise to other countries where political repression has been widespread.

The National Commission on the Disappeared, established by President Raúl Alfonsín in 1983, has estimated that more than 9,000 Argentines were illegally detained and killed under military rule, most of them between 1976 and 1978. Although many bodies were cremated or dumped at sea, others were buried in graves marked "N.N."—*ningún nombre* ("no name"). The forensic anthropology team, which was spawned by a workshop given in Buenos Aires by the American Association for the Advancement of Science, has so far exhumed some 120 bodies and has positively identified 30; some of the identifications have already served as evidence in trials that produced convictions.

The team gathers data by examining skeletal remains and also by studying records of prisoner transfers from illegal detention centers and burial records from cemeteries. Clyde C. Snow, an American forensic anthropologist based in Norman, Okla., taught the Argentine workers many of their techniques: how to determine age, sex, stature, handedness, race—even a woman's childbearing history—from fragmentary remains. Frequently, dental records or details of healed fractures will clinch an identification.

According to Mercedes Doretti, one of the team's founding members who visited Washington, D.C., in September, brain tissue has recently been found in skulls exhumed from a mass grave near Buenos Aires. The team hopes to extract DNA from the brain tissue or from the cranial bones to obtain DNA fingerprints—detailed analyses of the victims' DNA that can then be compared with the DNA of living relatives. By doing so, the team thinks it can identify the *desaparecidos* with a high degree of certainty. Cellmark Diagnostics in Germantown, Md., has offered to help by analyzing bone samples. Karen R. Rubenstein says the company is cautiously optimistic that it will be able to obtain DNA fingerprints from samples sent by the Argentine team. Immunologists at the Hospital Duran in Buenos Aires are already using specific genetic markers in the blood to identify children of *desaparecidos* who were kidnapped and secretly adopted by military families.

Laws passed in 1987 and 1988 by the Alfonsín government—which operated under the constant threat of military rebellion—reduced the number of trials of those who allegedly authorized the killings, from about 300 to 17. When President Menem announced his intention to pardon the accused, he also said he would consider granting pardons to some of the senior officers and junta leaders who were previously convicted. Nonetheless, the forensic anthropology team intends to continue its investigations, Doretti says, both for the sake of the families of the *desaparecidos* and for the historical record. Interest in the team's work is spreading: members have visited human-rights groups in the Philippines and Chile and are cur-

Most of the scores of exhumed Argentine victims are impossible for investigators to identify

REMAINS of an unidentified male, estimated to have been 40 to 45 years of age, were unearthed by the Argentine Forensic Anthropology Team. Six bullets were found in the skull and elsewhere in the skeleton. The official death certificate records the cause of death as "acute hemorrhage."

Before you buy a typewriter it pays to do your homework.

- [x] 7,000 Character Memory
- [x] 24 Character Display
- [x] Insert
- [x] Spell-Right® 50,000 Word Electronic Dictionary
- [x] Full-Line Correction
- [x] Correcting Cassette
- [x] Right Ribbon System™
- [x] WordFind®
- [x] List
- [x] WordEraser®
- [x] Auto Center/Return
- [x] Relocate
- [x] Auto Underscore
- [x] End of Page Warning
- [x] Dual Pitch
- [x] Bi-Directional Print
- [x] Stop Codes
- [x] Memory Battery Back-Up
- [x] Bold Print
- [x] Forward/Reverse Index
- [x] Auto Half-Space

Today's assignment is quite simple. And quite rewarding. Just study the remarkable features of the Smith Corona SD 750 and compare them with other typewriters.

After all, how many comparably priced type-writers give you word processing capabilities like Display and Memory, so you can edit, revise and make your work letter-perfect. And try and find the Spell-Right® 50,000 word Electronic Dictionary or the exclusive fumble-free Correcting Cassette on anything but a Smith Corona typewriter.

Though we've packed all these features into a portable that weighs under 14 pounds, we've been able to keep the cost equally lightweight.

The versatile Smith Corona SD 750. It makes buying a typewriter the easiest assignment you'll ever have.

SMITH CORONA®
TOMORROW'S TECHNOLOGY AT YOUR TOUCH™

For more information on this product, write to Smith Corona Corporation, 65 Locust Avenue, New Canaan, CT 06840 or Smith Corona Canada, 440 Tapscott Road, Scarborough, Ontario, Canada M1B 1Y4.

rently planning visits to Paraguay and Bolivia. —*T.M.B.*

Wires That Think

Philosophy and engineering mix at artificial-intelligence meeting

Notes from the 11th International Joint Conference on Artificial Intelligence (IJCAI): Detroit is a postindustrial city if ever there was one. The success of the automobile and the resulting hyperplasia of affluent suburbs have cut Motown's population in half; perhaps a fifth of the buildings that abut on the track of the downtown transit system are vacant.

Bringing 5,000 people here for IJCAI poses a conundrum: industrial cities like Detroit have declined while areas harboring service industries, such as computers, have grown. Nonetheless, applying artificial intelligence—that very highest of technologies—to the messy realities of the factory floor may be the only way for Detroit and indeed the entire U.S. to maintain a presence in traditional manufacturing.

Yet at the conference industrial concerns were conspicuous largely by their absence. A single panel—entitled "(How) Is AI Impacting Manufacturing?"—produced the observation that a little AI often goes a long way. Rather than spending time at IJCAI, industrial engineers, by most accounts, were elsewhere, attending conferences geared to aerospace or electronic design.

Researchers in attendance, meanwhile, discussed some of the more fundamental topics in AI:

Perception. Dana H. Ballard of the University of Rochester described a three-dimensional vision method that relies on parallax shifts caused by eye motion to extract qualitative information on the relative distances of objects in a scene. The technique mimics the physical constraints of animal depth perception, in contrast to conventional computer-vision methods, which attempt to eliminate eye motion and other such "extraneous" factors. Ballard's work meshed nicely with the discovery, presented elsewhere at the conference, that human infants develop an understanding of spatial relations among objects only as they begin to walk around and manipulate them.

Computer chess. An overflow crowd watched as Deep Thought, a special-purpose computer from Carnegie-Mellon University, soundly beat human grandmaster Robert Byrne, chess columnist for the *New York Times.* Byrne played on well past the point at which he would have been expected to capitulate to a human. Although chess computers are generally considered stronger on tactics than on strategy, one Canadian master commented that Deep Thought was giving Byrne "a little lesson in positional play." Although Deep Thought has been granted grandmaster ranking in the U.S., the World Chess Federation has refused to rank computer-chess programs or to permit them to compete in tournaments.

Integrated intelligence. The time is ripe, contended Allen Newell of Carnegie-Mellon, for combining the problem-solving, perception and learning strategies developed in separate AI projects and for bringing them to bear on large systems that will demonstrate not only knowledge but also autonomy. Robotics, natural-language understanding and other technologies are sufficiently robust, he said, to support intelligent systems that can explore the world and interact independently with human beings.

Symbolic and nonsymbolic reasoning. Most of the publicity in AI has been reserved for programs that use more or less strict, symbolic, logical procedures. Nevertheless, nonsymbolic systems such as neural networks, genetic algorithms (systems that combine and propagate the best of an initially random set of programs) and "bounded rationality" (a collection of methods for producing reasonable responses without working through a problem explicitly) all made appearances in Detroit. One workshop participant voiced surprise at seeing major proponents of both symbolic and nonsymbolic systems appearing in the same room, much less agreeing on many issues.

Planning. Intelligent systems that operate in the real world must be able to map out courses of action to accomplish their goals. Among the perennial issues discussed were methods for changing plans in midstream to accommodate new information as it becomes available, ways to recognize the plans of other entities with which the systems interact and techniques to decide how much time to spend formulating an efficient plan when time is of the essence.

The conference generated only two overflow crowds: one for the chess match, the other for the satellite link showing live pictures from *Voyager 2.* Nearly 2,000 attended an evening concert by the Detroit Symphony orchestra, but the gilt-and-wood Masonic Temple Theatre could have held twice the number. —*Paul Wallich*

BIOLOGICAL SCIENCES

Hungry to Evolve?

Can mistakes and malnutrition explain Lamarckian bacteria?

Conservative biologists shuddered over a year ago when they opened the pages of *Nature* and saw the specter of Jean-Baptiste Lamarck. In the September 8, 1988, issue, John Cairns and his colleagues at the Harvard School of Public Health presented evidence suggesting that under the right conditions cells can "choose" to mutate in beneficial and adaptive ways. The suggestion violated a central tenet of traditional genetics—that the frequency of a mutation is unrelated to its survival value—and smacked of Lamarck's previously discredited idea, first presented in 1809, that species evolve through the inheritance of acquired characteristics. Until recently drastic revisions in molecular biology appeared necessary to explain Cairns's results. Bernard D. Davis of Harvard Medical School, however, has offered a new explanation for nonrandom mutations that may save the established biological dogmas.

According to the widely accepted neo-Darwinian view, cells mutate continuously and randomly, and natural selection weeds out the least fit mutants. Cairns, however, suspected that this theory had never been adequately tested. He and his colleagues looked closely at a special strain of *Escherichia coli* bacteria that had lost the ability to digest the sugar lactose because of a defective gene. When the researchers grew the bacteria in lactose cultures, they found that lactose-digesting mutants cropped up improbably often. Some cryptic Lamarckian process seemed to be directing the mutations along adaptive lines.

More persuasive evidence for directed mutations emerged a few months later when Barry G. Hall, now at the University of Rochester, published data in *Genetics* showing that mutant bacteria capable of digesting a molecule called salicin appeared in cultures containing that substance. This ability could only have developed as the result of two specific mutation events, each extremely rare and neither conferring any advantages by itself. Thus, the odds of the salicin-digestion trait appearing by chance were vanishingly small.

As might be expected, these discoveries have touched off a controversy. Some biologists maintain that Cairns

WE DESIGN EVERY VOLVO TO LOOK LIKE THIS.

You're looking at a perfect Volvo. A Volvo that performed exactly as our safety engineers designed it to.

Its front and rear ends, for example, collapsed on impact. As a result, much of the crash energy was absorbed instead of being passed on to the passengers.

The car's middle section, however, didn't collapse. That's because the entire passenger compartment is surrounded by Volvo's unique "safety cage." Made of six box section steel pillars, this protective housing is strong enough to support the weight of six Volvos.

But the passengers of this car were also protected in ways you can't see. Because inside are such standard features as a driver's side Supplemental Restraint System, a collapsible steering column and, of course, 3-point seat belts, front and rear.

Every Volvo is designed to help protect its passengers in all these ways. And, as a result, will look remarkably similar to this one after being in the same type of accident.

If you're concerned about safety, you can't find a more beautiful car.

VOLVO
A car you can believe in.

CALL TOLL FREE FOR YOUR CATALOG OF 1990 VOLVOS: 1-800-221-9136.

and Hall must be mistaken. Others feverishly speculate that hitherto unknown mechanisms may direct beneficial mutations, possibly by creating copies of genes for proteins that prove serendipitously useful. The flow of information from proteins back to genes—an inversion of the normal DNA-to-RNA-to-protein route—contradicts key biological assumptions.

Now Davis has presented a less revolutionary alternative: that cells starved for nutrients may frequently make mistakes while transcribing their genetic information. Recently, in the *Proceedings of the National Academy of Sciences,* he suggested that hungry cells may cannibalize their own proteins for energy. If DNA-repair enzymes are among the proteins destroyed, mutations could become more probable. Because bacteria in nature are usually hungry, Davis suspects that malnutrition could be a powerful engine in driving their evolutionary change.

He also points out that when genes are transcribed to make proteins, DNA in the genes temporarily abandons its robust double-strand form and becomes single-stranded. Because single-strand DNA is much more vulnerable to damage and mutation, active genes may be more likely to mutate than idle ones. In the case of Cairns's experiments, Davis thinks that lactose caused parts of the bacteria's defective gene to separate into single strands, thereby creating a bias for mutations affecting lactose digestion.

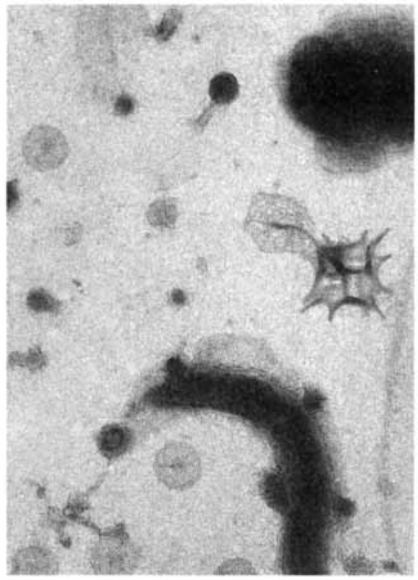
The debate over the origin of mutations rages on. Cairns insists that he has experimentally disproved Davis's theory; he also questions Davis's mechanism because it hinges on the unproved idea that genes are more mutable during translation. Alternative mechanisms continue to pop up. Will molecular biology resurrect Lamarck? Wait and see. —*John Rennie*

Economies of Scale

The smallest of the small swim in unsuspected numbers

So, naturalists observe, a flea/Hath smaller fleas that on him prey;/And these have smaller still to bite 'em;/ And so proceed ad infinitum.
—Jonathan Swift

Swift's verse might not be literally true, but recent findings suggest he was closer to the truth than he probably realized. The majority of aquatic microorganisms were once



VIRUSES cling to a banana-shaped bacterium in water taken from a Norwegian fjord. Electron micrograph by Mikal Heldal and Øivind Bergh.

thought to be nanoplankton, tiny organisms measuring no more than 10 or 20 micrometers in diameter. Then picoplankton, organisms less than two micrometers across, were found to be far more numerous than had been realized. Now it seems that the world's water is teeming with life's smallest forms: viruses no bigger than .2 micrometer in diameter.

Øivind Bergh, Mikal Heldal and their colleagues at the University of Bergen in Norway report in *Nature* that as many as 10 million—or even 100 million—viruses exist in a single milliliter of unpolluted water. That is about 10 times the number of bacteria found in the same milliliter and at least 1,000 times more than had been suspected. If the finding is correct, then viruses are almost certainly the most numerous life-forms on the earth. Does this suggest they play an important role in the ecology of aquatic ecosystems, or are they just innocent bystanders?

In water samples collected from fjords, lakes and the open ocean that were then analyzed under an electron microscope, the researchers found a population of mostly crystalline-shaped viruses. Yet a significant number had "tails," a trait that suggests they may be bacteriophages (viruses that attack bacteria). Indeed, some of the researchers' micrographs appear to show the viruses preying on bacteria. If so, then the viruses might account for the high turnover rates of bacteria in natural waters, comments Evelyn B. Sherr of the University of Georgia. Previously, ecologists had supposed that grazing by protozoa must be the principal cause of bacterial mortality in natural waters, but the researchers were perplexed because there were so few protozoa.

The large numbers of viruses in water could have other implications. The Bergen workers suggest the viruses might be an avenue for the transfer of genetic information among bacteria, including, presumably, any traits that human beings introduce deliberately by genetic engineering. Martin Alexander of Cornell University thinks it is premature to speculate about the role of viruses in genetic-information transfer. He suggests many of the particles might be accumulated remnants of long-dead viruses: "You could be looking at a graveyard," he says.

An alternative theory is that the viruses might help to explain the great diversity of bacteria in natural waters, a view held by Gunnar Bratbak, a member of the Bergen group. He speculates that viruses might preferentially attack whatever bacteria are most abundant, preventing a single bacterial species from becoming dominant and giving others a chance to proliferate.

Why has no one seen these viruses before? According to Bratbak, previous researchers examined viruses they had captured on filters. Once viruses are embedded in a filter, however, they cannot be examined by transmission electron microscopy, the technique that gives the highest magnification. Bratbak's group concentrated viruses with a centrifuge instead—and discovered a new realm. —*T.M.B.*

Sweet and Sour

The secret of sweetness is in the shape

The human species' insatiable craving for sweet-tasting foods must surely count as one of the great moving forces of history. Sugar cane was cultivated and refined in India more than 2,000 years ago; it later spread to the Mediterranean, probably with Arab invaders, and centuries later Columbus brought sugar production to the Caribbean. By 1550 sugar cane was the major export crop of Hispaniola (now Haiti and the Dominican Republic). Later, thousands of slaves were shipped from Africa to work the sugar plantations in order to satisfy New Englanders' taste for rum and molasses. Today the global market for sugar and artificial sweeteners is reckoned in tens of billions of dol-

lars. All the more surprising, then, that chemists have been quite unable to say what makes one substance sweet and another one bitter.

It has long been clear that sweetness is not just a simple chemical property like acidity, for example. Some sugars, such as sucrose (table sugar), are sweet, whereas other sugars, such as lactose, which is closely related to sucrose but of a different shape, are hardly sweet at all. In contrast, artificial sweeteners such as aspartame, a peptide consisting of the amino acids phenylalanine and aspartic acid, are chemically very different from sucrose. The form of aspartame used as a sweetener is 150 to 200 times sweeter than sucrose, yet differently shaped forms of aspartame are bitter or tasteless. Such observations suggest sweetness is related to the shape and size of a molecule, probably because the two properties determine how a molecule binds to taste-bud receptors in the mouth.

By having panels of judges taste dozens of molecular variants on a basic sweet-tasting shape, Murray Goodman of the University of California at San Diego and his colleagues seem to have refined the essence of sweetness in peptide-type sweeteners. They have recently encoded this knowledge in a computer program that, according to Goodman, can predict the sweetness or bitterness of a compound before it is synthesized.

Studies in a number of laboratories have shown that the molecules of many artificial sweeteners contain two ring structures, one of them polar (water-attracting) and one of them not. The two rings are connected by a short chain of atoms, which is itself polar and has a 90-degree bend in it. Because the rings lie nearly in the same plane, sweet molecules are roughly L-shaped. In aspartame, the aspartic acid provides the polar ring and the linking chain, and the phenylalanine provides the nonpolar ring; the molecule assumes its sweet-tasting shape when it is dissolved in water. When a sweet-tasting molecule is chemically modified so that the two rings no longer lie in the same plane, the effect is easily measured: the molecule becomes either bitter or tasteless.

Goodman has synthesized some molecules in which the connecting link between the rings is flexible; as a result, the rings can rotate with respect to one another. How do such molecules taste? Some taste bitter at first, then sweet; in others the sequence is reversed. Goodman suggests this occurs because at any moment some molecules are in a sweet conformation and so will bind to sweetness receptors, whereas other molecules are in a bitter conformation and will bind to distinct bitterness receptors. Chemical modifications probably affect the binding efficiency of each conformation, thus influencing which taste emerges first. Molecules that lack the 90-degree bend in the connecting link or that have too large a nonpolar ring generally have no taste at all—presumably because such molecules cannot bind to receptors.

Goodman says his computer program can account for the sweetness of compounds such as saccharin, cyclamate, acesulfame-K (Sunette) and alitame, an artificial sweetener under development by Pfizer, Inc. New low-calorie synthetic sweeteners designed with the aid of molecular modeling are in the pipeline, Goodman says. But the molecules that have so far eluded analysis are the sugars: sucrose, lactose, fructose and so on. These molecules are apparently too flexible for Goodman to know their shape precisely. In addition, he notes, they are not even all that sweet as compared with the peptide-type sweeteners. Sweetness, it seems, still has some secrets. *—T.M.B.*

PHYSICAL SCIENCES

Siberian Snake

A Russian-born technique may improve proton accelerators

It is not some Asian adder, nor a communist con artist. The Siberian snake is a magnetic contraption that keeps protons spinning in unison as they whirl around an accelerator. Recent experiments suggest the device could greatly enhance the ability of proton colliders—including the proposed Superconducting Supercollider (SSC)—to test fundamental theories of matter.

The property known as spin is exhibited by all subatomic particles and plays a crucial role in their interactions. Determining how spin affects particles smashed together in an accelerator is much easier if the particles are polarized, their spins aligned rather than randomly oriented. Unfortunately protons, the particles of choice for many of the world's leading colliders, become ever more difficult to keep polarized as they are boosted to higher energies. Physicists have devised complicated methods for overcoming the problem in accelerators such as the Alternating Gradient Synchrotron (AGS) at the Brookhaven National Laboratory, but for higher-energy machines these approaches are impractical.

In 1974 two Soviet physicists, Yaroslav S. Derbenev and Anatoly M. Kondratenko of the Novosibirsk Laboratory, suggested that passing proton beams through an undulating magnetic field might counteract depolarizing influences. The proposal intrigued many physicists (including Ernest D. Courant of Brookhaven, who is credited with naming it the "Siberian snake"), but none of the world's accelerators seemed to have the space for the lengthy battery of magnets needed to test the idea.

This past summer, however, researchers at the University of Michigan, Brookhaven and Indiana University wrapped a 10-meter-long snake around an accelerator at Indiana and tested it—with successful results. Alan D. Krisch of Michigan, who led the experiment, hopes to test the snake next on a larger machine, such as the Fermi National Accelerator Laboratory's Tevatron. If this experiment is also successful, Krisch says, the designers of several powerful accelerators now being planned—including the SSC, West Germany's HERA and the Soviet Union's UNK—may consider adding the snake to their designs.

If these machines could generate polarized protons, notes Lazarus G. Ratner of Brookhaven, they might help answer questions raised recently about quantum chromodynamics, currently the best theory describing the strong nuclear force binding protons and neutrons together. Doubts about the theory, Ratner says, have grown out of polarized-proton experiments done in Brookhaven's AGS and other colliders [see "Collisions between Spinning Protons," by Alan D. Krisch; SCIENTIFIC AMERICAN, August, 1987].

Snakes need not be terribly expensive, according to Lee C. Teng of Fermilab and the Argonne National Laboratory. He estimates that equipping the Tevatron with a snake would cost about $20 million and the SSC about twice that amount. Teng warns, however, that institutional inertia—together with the feeling of many particle physicists that spin is not a vital parameter in high-energy interactions—may prevent the device from being implemented in the Tevatron or the SSC.

SSC officials, still struggling to maintain congressional support and fund-

SCIENCE/SCOPE®

Technology which allows small satellite earth stations to transmit and receive voice, video, and data information in complete privacy helps smaller companies enjoy the advantages of satellite communication services. Hughes Aircraft Company has established and is operating earth station facilities which can be shared among many users as the central control point for their independent networks. When combined with Very Small Aperture Terminal (VSAT) stations using advanced transmission techniques, Hughes' shared hub facilities allow companies to quickly and cost-effectively establish their own private, customized, satellite-based business communications networks. The first three Hughes shared hubs are located in New York, Los Angeles and Minneapolis; additional sites will be added in the future.

Electro-optic sensors will track targets even in highly cluttered environments. The Dual Mode Tracker, developed and being readied for production by Hughes, has the capability to track missiles, aircraft and vehicles. Using an advanced correlation/centroid tracker, the system can simultaneously track targets while automatically selecting the preferred mode of tracking. It will also automatically acquire moving targets in the field of view of any video imaging sensor connected to it. The Dual Mode Tracker can be utilized in tactical and strategic applications.

Microwave transmissions of more than 80 television channels over extended distances have been made possible by a new solid state broadband transmitter, developed by Hughes, with the highest power output in its frequency range of 13 Gigahertz. The indoor transmitter, designated the Model IBBT-116, uses the latest power doubling and feed forward gallium arsenide technology to achieve a 6-fold increase in output performance. Besides increasing the distance over which signals can be transmitted, the new equipment permits more receiving points than similar transmitters. In addition to cable TV and other kinds of microwave signals, the new technology can be used in satellite and ground terminal applications.

A new program will enhance the capabilities of F-4 Phantom fighters flying for the air forces of Egypt, Greece, Turkey, and South Korea. Hughes is teamed with McDonnell Douglas Corp. and Messerschmitt-Boelkow-Blohm (MBB) to market an F-4 Phantom upgrade which includes putting Hughes APG-65 radars into approximately 250 aircraft. The APG-65 will provide the Phantoms with digital, multimode operation providing look-down, shoot-down capability, in addition to the radar's multi-track, air-to-air capability. The radar can also launch the fire-and-forget Advanced Medium Range Air-to-Air Missile (AMRAAM), Sparrow and Sidewinder missiles in the air-to-air role and the air-to-ground weapons in the current NATO inventory.

Hughes' Combat Systems Engineering Facility in San Diego, California has immediate openings in advanced development and training to support the Navy Command and Control Processor (C2P) and Advanced Combat Direction System (ACDS) Programs. Experience desired for Combat Systems Engineers includes 7-9 years of system level development of military systems, preferably Surface Navy Combat Systems. For Computer Programmers/Instructors the level of experience desired is 4-5 years of designing, coding and debugging computer software. Teaching or training experience is desired. Applicants must have a B.S. Degree in Computer Science or the equivalent. Please send your resume to Hughes Aircraft Company, Ground Systems Group, Dept. S2, P.O. Box 4275, Fullerton, CA 92634. Equal opportunity employer. U.S. citizenship required.

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

HUGHES

Subsidiary of GM Hughes Electronics

ing for their project, acknowledge that the snake issue is not at the top of their list of priorities. But Donald A. Edwards of the SSC's central design group says the collider's managers intend to examine the snake's potential soon. Even if the managers decide not to implement the technique initially, he notes, they may well leave room for it to be added later. *—J.H.*

Hard-Pressed

Squeezed hydrogen forms metal with superconducting potential

Stupendous pressure generated in a nutcrackerlike device has apparently turned hydrogen into metal. The conversion, if true, may help confirm the theory of superconductivity; in addition, it may provide planetary scientists with insights into the structures of Jupiter, Saturn and the other giant planets, whose interiors are thought to consist largely of metallic hydrogen.

Ho-Kwang Mao and Russell J. Hemley of the Carnegie Institution in Washington, D.C., describe their success in *Science*. The researchers subjected a tiny sample of hydrogen, in an area less than 20 microns wide (one one-hundredth the diameter of a human hair), to a steady pressure of more than 2.5 million atmospheres in a press called a diamond-anvil cell [see "The Diamond-Anvil High-Pressure Cell," by A. Jayaraman; SCIENTIFIC AMERICAN, April, 1984]. This apparatus uses a system of screws and levers to squeeze a sample between two beveled gem-quality diamonds, the only material that is both strong enough to bear such record pressures and transparent enough to provide a window on the hydrogen sample. The diamonds must be perfectly aligned, or they will shatter—at a cost of several thousand dollars a crack.

Although the researchers had to run dozens of tests over a span of several days, the transformation of the hydrogen takes only two or three minutes and can be seen through a microscope. At first the hydrogen is squeezed to a transparent solid form that is known to be nonconducting. But above two million atmospheres the solid gradually darkens, indicating its transition to a semiconductor. In regions of the sample where the pressure approaches an estimated three million atmospheres, the hydrogen becomes opaque and scatters light in patterns that seem to indicate the absence of a molecular structure. The disappearance of molecules is consistent with the formation of a metal.

The study of hydrogen metal should test the theory of the solid state, because hydrogen's simple atomic structure lends itself to precise theoretical predictions. For example, it is calculated that hydrogen metal will superconduct at temperatures above the freezing point of carbon dioxide (−79° C), which is higher than even those reported recently for certain ceramics.

But can hydrogen metal retain its form when the pressure is withdrawn? Although Mao and Hemley found that their hydrogen sample became transparent again after they released the pressure, Mao thinks a gentler decompression may preserve the opaque, metallic form. Still, that would require a more delicate touch than anyone can yet muster. "Breakthroughs are hard to predict, but I would say it will take about a decade," Mao says.

Much hangs on whether hydrogen metal can be brought into the thin air of everyday life. Not only does the metal promise to conduct electricity with zero resistance near room temperature, but its unequaled energy content, many times that of liquid hydrogen, might make it an ideal rocket fuel or high explosive. Larger presses may one day crunch out tiny ingots of hydrogen, no doubt at great expense. But at this early stage the only certainty is that metallic hydrogen will be cheaper to produce on the earth than to extract from the heart of Jupiter. *—Philip Ross and Russell Ruthen*

Just a Veneer

The upper crust of continents can slip and deform on its own

Simple versions of plate-tectonic theory present the earth's crust as a passenger riding piggyback on 100-kilometer-thick plates of lithosphere. In regions where tectonic processes are reshaping the continents, however, the picture gets more complex: a network of faults breaks the landscape into fragments that seem to set their own courses. It now appears that large tracts of surface crust in such regions may actually be physically independent of the underlying material, slipping and deforming in a layer only five or 10 kilometers thick.

Evidence for such an independent upper crust is presented in *Geology* by B. Clark Burchfiel, Peter Molnar and their colleagues at the Massachusetts Institute of Technology and workers at seismological bureaus in Beijing and Yinchuan, China. It comes from detailed geologic mapping in two regions where tectonic stresses are dramatically deforming the continental crust: the Basin and Range province of the western U.S. and Tibet. In each region the authors studied structures along a strike-slip fault, a boundary at which blocks of crust slide past each other.

Where a strike-slip fault ends, the slip must stop and the relative motion along the fault must somehow be absorbed. The block trying to push past the fault tip may slide upward along a sloping transverse rupture called a thrust fault; the block pulling away from the tip may slip downward along a thrust fault's converse, a normal fault. By identifying and mapping the associated surface structures, one can infer how deep the faults extend and so how thick the layer in motion is.

At each end of the Hunter Mountain fault, a 50 kilometer strike-slip fault near Death Valley in California, extension across normal faults has opened a basin. If the normal faults plunged deep into the crust, the trailing edge of each moving block would have dropped sharply, generating dramatic relief. Yet the group found that even though the basins have accommodated the full 10 kilometers or so of relative motion along the Hunter Mountain fault over geologic time, they are only a few hundred meters deep. Based on the basins' proportions and shapes, the authors infer that each normal fault slopes gradually to a depth of no more than five kilometers, then flattens into a horizontal "detachment surface" underlying the moving crust.

In Tibet the team mapped thrust faults produced by compression at one end of the Haiyuan strike-slip fault, at the northeast corner of the Tibetan plateau. Although the structures have absorbed as much as 15 kilometers of horizontal motion over the past two million years, they record little uplift. The underlying thrust faults seem to reach no deeper than five or six kilometers—to what is presumably the base of the moving crust.

The authors think continental crust may behave like a thin skin over much larger areas than they have mapped directly. Slip along one side of the 1,000-kilometer Altyn Tagh fault of Tibet is absorbed in a zone of mountains and valleys produced by folds and shallow thrust faults. The faults seem to merge with a detachment surface underlying a broad expanse of moving upper crust.

How does such a veneer of crust come unglued from the rest of the plate? The investigators point out that crustal rocks soften at the high temperatures experienced below a depth of about 10 kilometers. The resulting zone of weakness might tend to free the upper crust from the deepest parts of the plate, which are made up of mantle rocks that remain strong in spite of high temperatures.

Are surface motions, then, entirely independent of plate motions in regions of continental deformation? Molnar thinks not: "The surface probably gives us a weighted average of what's going on at depth." Now, he says, geologists must ask how tectonic forces could be communicated to this thin crustal skin. —*Tim Appenzeller*

Stillborn

Is an unusual "protogalaxy" really just an ordinary dwarf?

Late this past summer two astronomers from Cornell University discovered what seemed to be an unborn galaxy: a vast cloud of hydrogen gas in which no stars had yet formed. In a press release, Riccardo Giovanelli and Martha P. Haynes said their finding—made with the 300-meter radiotelescope at Arecibo, Puerto Rico—contradicted the presumption of many astronomers that all galaxies formed shortly after the birth of the universe. The *New York Times* thought the story fit for page one. The headline proclaimed: "Birth of Galaxy Seen in Gas Cloud: Discovery Seems to Undercut Classic Theory on Process."

Does it? Giovanelli and Haynes concluded that the cloud was a "protogalaxy" based on their inability to detect starlight in photographic plates made by the Mount Palomar Telescope in California. But workers at the University of Cambridge examined plates made by a more sensitive telescope in Australia and found a cluster of stars in the direction of the cloud. The Cambridge workers, Richard McMahon and Mike Irwin, think the object found by Giovanelli and Haynes is not a protogalaxy but a dwarf galaxy imbedded in a large envelope of gas—odd, perhaps, but far from unique.

"I disagree," Giovanelli retorts. He maintains if the starlight does come from the hydrogen cloud, it is too faint to represent even a dwarf galaxy. It is also possible, he says, that the starlight comes from an object in front of or behind the cloud. "A one-in-a-million chance," McMahon replies. Despite their differences, the two parties have agreed to collaborate in an effort to resolve the issue. —*J.H.*

MEDICINE

Winning Candidate

A painstaking search identifies the gene for cystic fibrosis

After a two-year molecular marathon along a fragment of a human chromosome, investigators at the Hospital for Sick Children in Toronto and the Howard Hughes Medical Institute at the University of Michigan have found and described the gene that causes cystic fibrosis, one of the most common fatal genetic diseases. The achievement represents a huge step forward in understanding of the tragic disease and opens the way to the development of therapies.

Cystic fibrosis (CF) is most prevalent in Caucasian populations, where it affects one in 2,000 live births. The disease involves the lungs, the digestive tract and sweat glands; it is the abnormally thick mucus produced in the lungs, which clogs airways and renders patients prone to infections, that kills most CF patients before the age of 30. Genetic studies showed years ago that the gene causing the disease is recessive: a person must inherit two defective genes, one from each parent, in order to acquire the condition. Until now, however, little has been known about the biochemical defect that underlies the symptoms.

Lap-Chee Tsui and John R. Riordan of the Hospital for Sick Children and Francis S. Collins of the University of Michigan led the teams that collaborated to identify the CF gene. Their findings, published in a series of papers in *Science,* end a highly competitive race for the gene. In 1987 workers at St. Mary's Hospital Medical School in London thought they had found a "candidate" gene, but it turned out to be only a close neighbor of the actual CF gene. The new claim is on firmer ground: the investigators have found not only a candidate gene but also a mutation in it that is present in 70 percent of a large sample of CF patients but in none of a large group of people who do not have CF. The researchers think such findings constitute proof that their gene is the one responsible for CF.

The gene, it turns out, is spread over a long stretch of DNA (250,000 base pairs) on chromosome 7. Most of that


CASIO SCIENTIFIC CALCULATOR
fx-7000G GRAPHICS
X=0.1063829787
TEXAS INSTRUMENTS TI-30SL
SHARP
SHARP
Scientific calculator
EL-506P


CASIO
X=-0.670212766

stretch, however, consists of regions (introns) that do not code for protein. The gene actually consists of 24 separate fragments (exons). Together they encode a protein, composed of some 1,480 amino acids, that has been named the cystic fibrosis transmembrane conductance regulator (CFTR). The protein's shape (deduced from its amino acid sequence) suggests it spans the cell membrane. The ability of the cells of CF patients to control the flow of ions across their membranes seems to be impaired, and so CFTR may be part of a pore that passes ions into and out of cells—or perhaps a molecule that controls such a pore.

CFTR has a region that binds adenosine triphosphate (ATP), the cell's energy currency, which suggests that the protein plays an active role in ion transport. The mutation found in 70 percent of CF patients causes the protein to lack one amino acid, a phenylalanine that normally lies in the ATP-binding site; the lack of the phenylalanine might somehow hinder ATP binding and so cripple the protein.

The investigators tracked down the CF gene with a battery of techniques. They and other workers had previously established that the gene lay on chromosome 7, close to two genetic landmarks. The Toronto and Michigan teams combined genetic analysis of DNA from affected families with techniques known as chromosome walking and jumping, which enabled them to clone, or generate in large quantities, dozens of identifiable fragments of DNA lying farther and farther away from the landmarks. Eventually they identified a gene, at a site implicated by genetic studies, that carried a mutation in most cases of CF. Both teams are now seeking to identify other mutations in the gene that are responsible for the remaining cases.

One immediate application of the new knowledge is likely to be the development of tests to identify CF carriers directly. Because as many as one in 20 Caucasians are CF carriers, there will be a large market for such tests, and commercial interest is intense. New therapies for CF sufferers are a longer-term possibility. The investigators are now trying to produce mice with CF by finding the equivalent mouse gene (if there is one) and inducing mutations in it. With that tool in hand, and with more detailed knowledge of the structure of CFTR, it may become feasible to design and test drugs or other therapies to correct the defect. *—T.M.B.*

Myocardial Infraction?

The evidence on cholesterol is subjected to interpretation

It started with a simple correlation: middle-aged men with high levels of cholesterol in their blood serum have more heart attacks than those with low levels of serum cholesterol. The correlation makes at least superficial sense, since it is cholesterol deposited in the coronary arteries that leads to heart attacks. But while many Americans dutifully eschew cheese omelettes and red meat in efforts to lower their cholesterol counts, some doctors are debating the impact of such efforts on myocardial health.

The latest installment in the debate appears in the *New England Journal of Medicine* in a pair of articles by two Boston physicians, Allan S. Brett of the New England Deaconess Hospital and Alexander Leaf of the Massachusetts General Hospital. Brett's article sets out to examine the "language and emphasis" of reports on two seminal cholesterol-intervention studies known as the Lipid Research Clinics (LRC) trial and the Helsinki Heart Study trial. While Brett says trial investigators have a "prerogative" to cast their findings "in the most optimistic light," his piece amounts to an indictment of their interpretations. In a rejoinder, Leaf maintains that interventions—whether through diet or drug therapy—are advisable nonetheless.

Brett claims that the researchers "colored" their study results by expressing outcomes in terms of relative reductions in the incidence of cardiovascular events rather than in absolute numbers. Thus, investigators in the LRC trial reported a 19 percent reduction in coronary deaths and nonfatal heart attacks among men treated with the cholesterol-lowering drug cholestyramine, when in absolute numbers there was only a 1.7 percent difference between the two groups: from 9.8 percent of the untreated group (187 of 1,900) to 8.1 percent of the treated group (155 of 1,906). Similarly, workers in the Helsinki study, which used the drug gemfibrozil, described an absolute difference of 1.4 percent in coronary disease as a 34 percent relative reduction.

"This descriptive language, although technically accurate, may operate subliminally to magnify the effect of the intervention for the reader," Brett writes. Leaf acknowledges the underwhelming success of intervention in the LRC trial but defends its utility: "What such studies have shown is that even slight decreases in cholesterol levels, when sustained over several years, can make a difference." And Leaf denounces the reliance on coronary death as an exclusive index in intervention studies, saying, in essence, that good health is more than the absence of death.

William P. Castelli, director of the 40-year-old Framingham Heart Study, says that the effects of intervention are becoming more obvious as cholesterol studies themselves reach middle age. Brett does not disagree. But, he adds, the public should be aware that for now some of the rationale for controlling serum cholesterol is based on extrapolations from existing clinical results. *—Karen Wright*

PROFILE

Quantum Consciousness

Polymath Roger Penrose takes on the ultimate mystery

Roger Penrose is slight in figure and gentle in mien, and he is an oddly diffident chauffeur for a man who has just proposed how the entire universe—including the enig-

ma of human consciousness—might work. Navigating from the airport outside Syracuse, N.Y., to the city's university, he brakes at nearly every crossroad, squinting at signs as if they bore alien runes. Which way is the right way? he wonders, apologizing to me for his indecision. He seems mired in mysteries.

When we finally reach his office, Penrose finds a can labeled "Superstring" on a table. He chuckles. On the subject of superstrings—not the filaments of foam squirted from this novelty item but the unimaginably minuscule particles that some theorists think may underlie all matter—his mind is clear: he finds them too ungainly, inelegant. "It's just not the way I'd expect the answer to be," he observes in his mild British accent.

When Penrose says "the answer," one envisions the words in capital letters. He confesses to agreeing with Plato that the truth is embodied in mathematics and exists "out there," independent of the physical world and even of human thought. Scientists do not invent the truth—they discover it. A genuine discovery should do more than merely conform to the facts: it should *feel* right, it should be beautiful. In this sense, Penrose feels somewhat akin to Einstein, who judged the validity of propositions about the world by asking: Is that the way God would have done it? "Aesthetic qualities are important in science," Penrose remarks, "and necessary, I think, for great science."

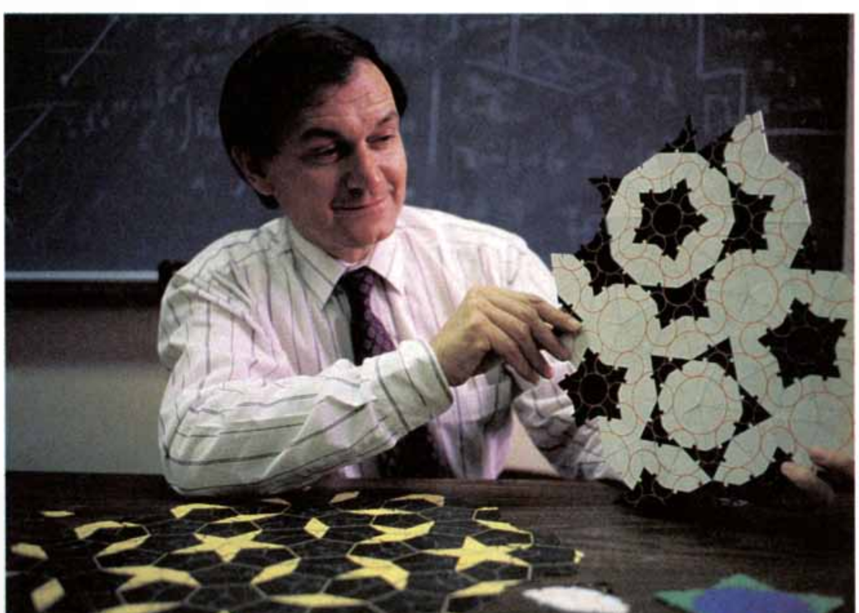
I interviewed Penrose in September while he was visiting Syracuse University, on leave from his full-time post at the University of Oxford. At 58 he is one of the world's most eminent mathematicians and/or physicists (he cannot decide which category he prefers). He is a "master," says the distinguished physicist John A. Wheeler of Princeton University, at exploiting "the magnificent power of mathematics to reach into everything."

An achievement in astrophysics first brought Penrose fame. In the 1960's he collaborated with Stephen W. Hawking of the University of Cambridge in showing that singularities—objects so crushed by their own weight that they become infinitely dense, beyond the ken of classical physics—are not only possible but inevitable under many circumstances. This work helped to push black holes from the outer limits of astrophysics to the center.

In the 1970's Penrose's lifelong passion for geometric puzzles yielded a bonus. He found that as few as two geometric shapes, put together in jigsaw-puzzle fashion, can cover a flat surface in patterns that never repeat themselves. "To a small extent I was thinking about how simple structures can force complicated arrangements," Penrose says, "but mainly I was doing it for fun." Called Penrose tiles, the shapes were initially considered a curiosity unrelated to natural phenomena. Then in 1984 a researcher at the National Bureau of Standards discovered a substance whose molecular structure resembles Penrose tiles. This novel form of solid matter, called quasicrystals, has become a major focus of materials research [see "Quasicrystals," by David R. Nelson; SCIENTIFIC AMERICAN, August, 1986].

Quasicrystals, singularities and almost every other oddity Penrose has puzzled over figure into his current magnum opus, *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics,* published this month by the Oxford University Press. The book's ostensible purpose is to refute the view held by some artificial-intelligence enthusiasts that computers will someday do all that human brains can do—and more. The reader soon realizes, however, that Penrose's larger goal is to point the way to a grand synthesis of classical physics, quantum physics and even neuropsychology.

Penrose thinks quantum gravity may underlie such puzzles as quasicrystals and consciousness



ROGER PENROSE holds up a sheet of Penrose tiles, geometric shapes that generate nonperiodic patterns. The photograph was made by Stephen Sartori.

He begins his argument by slighting computers' ability to mimic the thoughts of a mathematician. At first glance, computers might seem perfectly suited to this endeavor; after all, they were created to calculate. But Penrose points out that Alan M. Turing himself, the original champion of artificial intelligence, demonstrated that many mathematical problems are not susceptible to algorithmic analysis and resolution. The bounds of computability, Penrose says, are related to Gödel's theorem, which holds that any mathematical system always contains self-evident truths that cannot be formally proved by the system's initial axioms. The human mind can comprehend these truths, but a rule-bound computer cannot.

In what sense, then, is the mind unlike a computer? Penrose thinks the answer might have something to do with quantum physics. A system at the quantum level (a group of hydrogen atoms, for instance) does not have a single course of behavior, or state, but a number of different possible states that are somehow "superposed" on one another. When a physicist measures the system, however, all the superposed states collapse into a single state; only one of all the possibilities seems to have occurred. Penrose finds this apparent dependence of quantum physics on human observation—as well as its incompatibility with macroscopic events—profoundly unsatisfying. If the quantum view of reality is absolutely true, he suggests, we should see not a single cricket ball resting on a lawn but a blur of many balls on many lawns.

He proposes that a force now conspicuously absent in quantum physics—namely gravity—may link the quantum realm to the classical, deterministic world we humans inhabit. That idea in itself is not new: many theorists—including those trying to weave reality out of superstrings—have sought a theory of quantum gravity. But Penrose takes a new approach. He notes that as the various superposed states of a quantum-level system evolve over time, the distribution of matter and energy within them begins to diverge. At some level—intermediate between the quantum and classical realms—the differences between the superposed states become gravitationally significant; the states then collapse into the single state that physicists can measure. Seen this way, it is the gravitational influence of the measuring apparatus—and not the abstract presence of an observer—that causes the superposed states to collapse.

Penrosian quantum gravity can also help account for what are known as nonlocal effects, in which events in one region affect events in another simultaneously. The famous Einstein-Podolsky-Rosen thought experiment first indicated how nonlocality could occur: if a decaying particle simultaneously emits two photons in opposite directions, then measuring the spin of one photon instantaneously "fixes" the spin of the other, even if it is light-years away.

Penrose thinks quasicrystals may involve nonlocal effects as well. Ordinary crystals, he explains, grow serially, one atom at a time, but the complexity of quasicrystals suggests a more global phenomenon; each atom seems to sense what a number of other atoms are doing as they fall into place in concert. This process resembles that required for laying down Penrose tiles; the proper placement of one tile often depends on the positioning of other tiles that are several tiles removed.

What does all this have to do with consciousness? Penrose proposes that the physiological process underlying a given thought may initially involve a number of superposed quantum states, each of which performs a calculation of sorts. When the differences in the distribution of mass and energy between the states reach a gravitationally significant level, the states collapse into a single state, causing measurable and possibly nonlocal changes in the neural structure of the brain. This physical event correlates with a mental one: the comprehension of a mathematical theorem, say, or the decision not to tip a waiter. The important thing to remember, Penrose says, is that this quantum process cannot be replicated by any computer now conceived.

With apparently genuine humility, Penrose emphasizes that these ideas should not be called theories yet: he prefers the word "suggestions." But throughout his conversation and writings, he seems to imply that someday humans (not computers) will discover the ultimate answer—to everything. Does he really believe that? Penrose mulls the question over for a moment. "I guess I rather do," he says finally, "although perhaps that's being too pessimistic." Why pessimistic? Isn't that the hope of science? "Solving mysteries, or trying to solve them, is wonderful," he replies, "and if they were all solved that would be rather boring." *—John Horgan*

Why it's vital to all Amer environmental trust fu

Somebody's cleaning up, but it's not the environment.

Federal legislation creating Superfund was first enacted almost 10 years ago.

Priority action has been set for 1,200 of the most hazardous waste sites. And the list is growing.

Only 27 have been cleaned up.

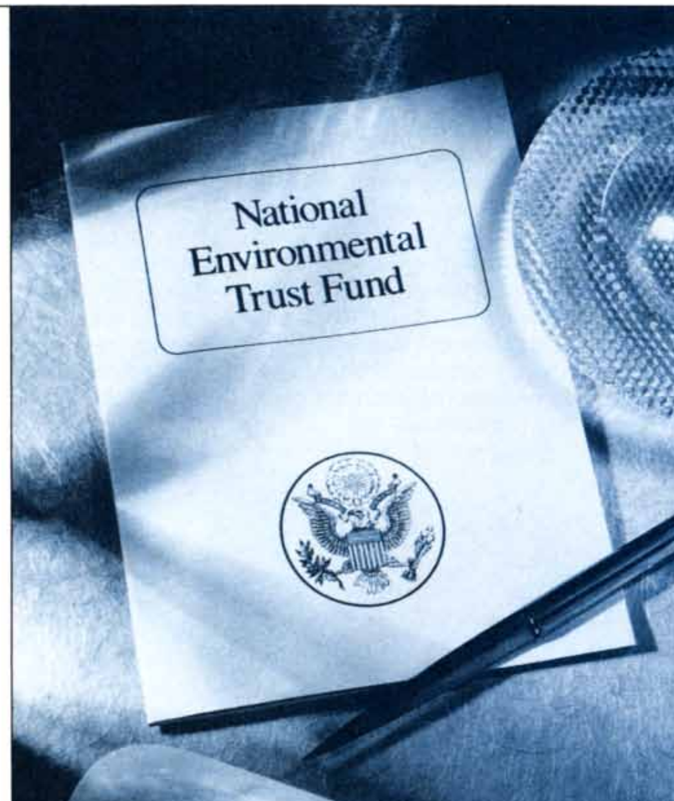
Despite the fact that at many sites it's impossible to tell who caused what and when, as much as 60% of cleanup funds go toward legal expenses in costly efforts to fix the blame instead of the problem.

Why should one party bear the sole burden for what amounts to a broad public problem?

Businesses and public agencies are being held liable today for environmental damage that was not considered illegal or intentional 25, 30 or 40 years ago.

The cost to clean up just one existing toxic waste site can run as high as several billion dollars. If individual companies are found responsible, the resulting expense could put them at a serious competitive disadvantage.

Why? Because current law requires them to pay for cleaning up the residue of economic progress–progress from which we've all benefitted.

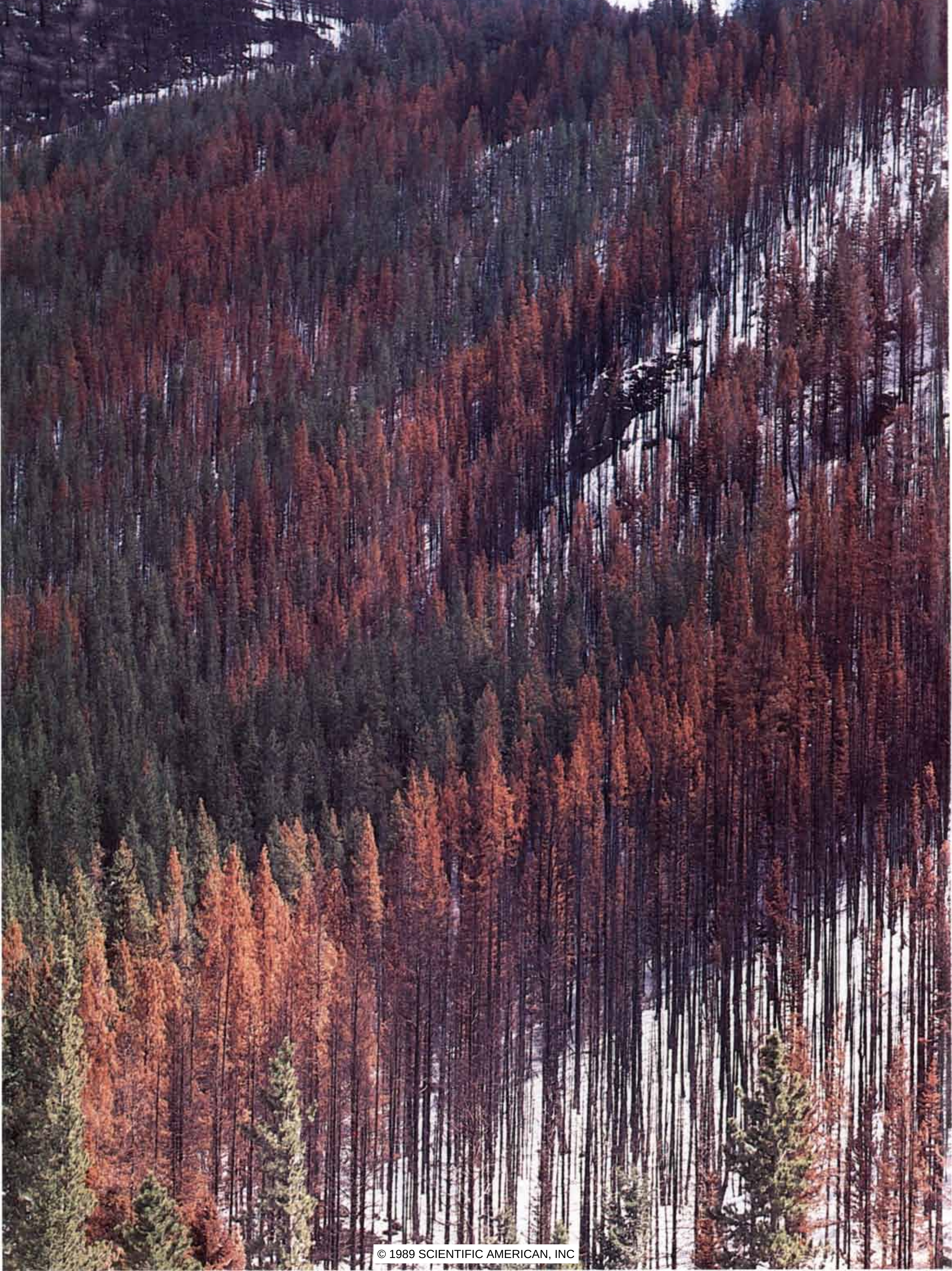


Legislation is needed to create a National Environmental Trust Fund.

There's more action in the courtroom than at the clean up sites.

Imagine holding a construction company liable for rebuilding an old office building whenever a new building code is issued. This form of retroactive liability is doing as much harm to our efforts to deal with toxic waste as toxic waste has done to the environment.

At AIG, we think it's high time to find a



SCIENTIFIC AMERICAN November 1989 Volume 261 Number 5

The Yellowstone Fires

During the summer of 1988 fires swept across much of Yellowstone National Park. Why did so many acres burn in one year? A look at the ecological history of the region provides some answers

by William H. Romme and Don G. Despain

The summer of 1988 was an eventful one in Yellowstone National Park. Fires that began in June and July were raging across the landscape by August. On August 20, a day that later became known as Black Saturday, more square miles burned in a single 24-hour period than had burned during any decade since 1872. The situation was beyond control; even with the best available technology—and extraordinary skill and courage on the part of the fire fighters—the fires kept spreading. It was not until the arrival of snow in mid-September that the fires finally lost their strength, although they did not die out completely until the onset of winter in November.

Having been highly criticized for not extinguishing the fires when they began (although existing policy was to let fires burn under certain conditions), officials of Yellowstone Park are now reevaluating fire management in natural ecosystems. It is not surprising that the unusual events of 1988 raise many questions about the natural history of fires and their role as forces of ecological change.

What accounted for the extraordinary severity of the Yellowstone fires? Why were they so large and so hard to control? Could they have been prevented by human intervention? Should they have been prevented? The fires also raise important questions about the kind of fire management that is needed in Yellowstone. A review of the summer's chronology provides a vantage from which such concerns can be addressed.

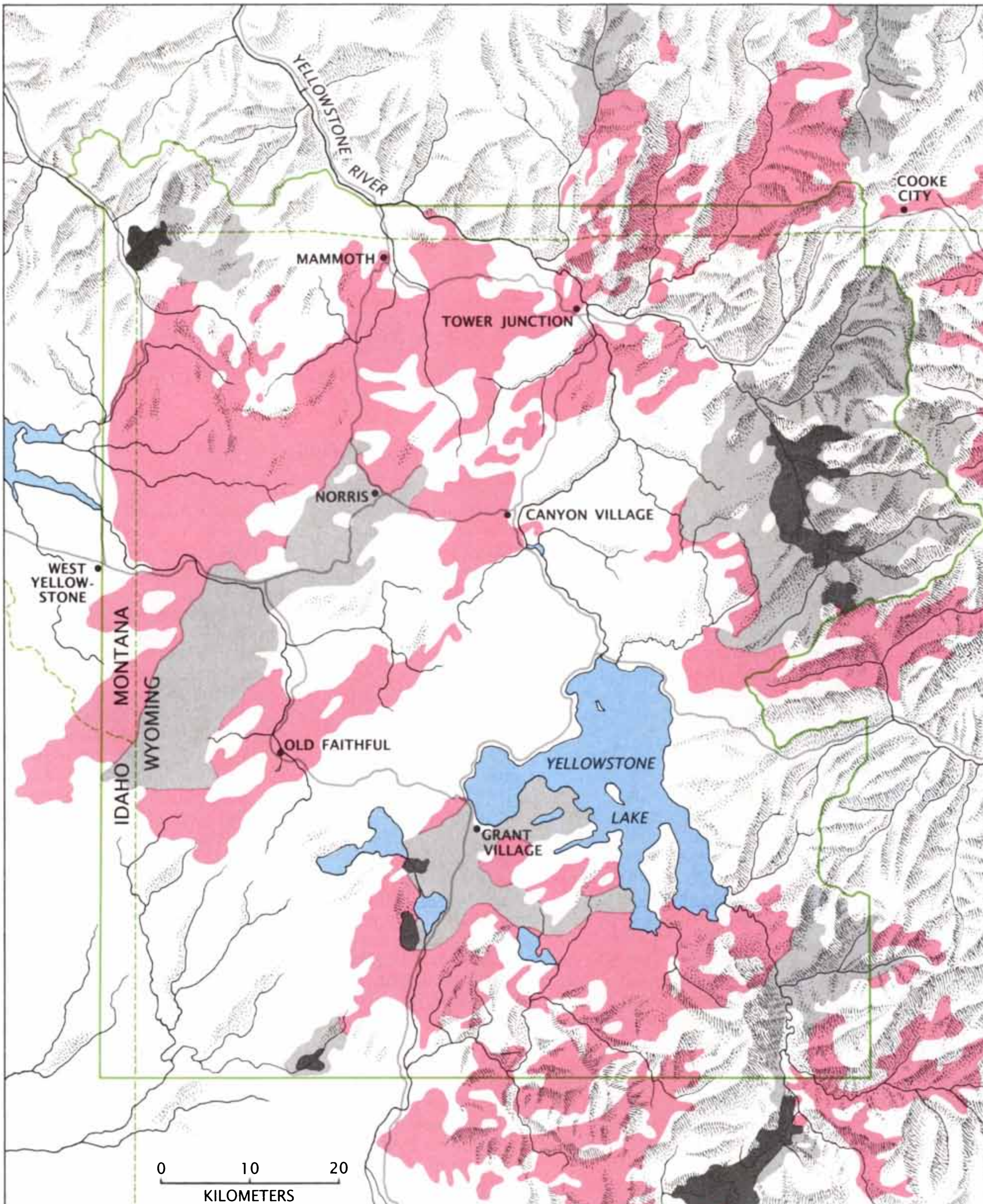
The first fires in the greater Yellowstone ecosystem, a region that includes Yellowstone National Park and adjacent lands, were ignited by lightning. The previous winter had been dry, as had all the winters since 1982, but the previous six summers had been wetter than average, and so it appeared that 1988 would follow the same pattern. Indeed, when lightning ignited several fires in June there was no reason to think that human life, property or significant resources would be threatened, and the fires were allowed to burn without interference according to an approved management plan.

But the usual rains of June and July failed to materialize, and by July 15 some 8,600 acres had burned. At that time a decision was made to suppress, or extinguish, a large fire that threatened to cross the park's southern boundary. By July 21 fires had burned across more than 17,000 acres, and it was evident that action would have to be taken to suppress the ones that were still burning. On that day Yellowstone officials also decided to suppress all new fires, regardless of whether they were caused by lightning or by human beings. By August 21 the number of acres burned had risen to some 400,000. Why, despite massive suppression attempts, did the fires continue to spread?

Weather, it seems, was a major factor. Winds as high as 100 miles per hour combined with high temperatures and droughtlike conditions (the summer of 1988 was the driest since the dust bowl days of the 1930's) to create impossible odds for the more than 9,000 fire fighters who were brought in to battle the blaze. Adding to their difficulties was the self-sustaining momentum that fires acquire as they get bigger. As more and more heat is generated, the volume of hot air in the immediate vicinity of a fire

FIRES VARY in intensity as they cross a landscape. In some areas they burn everything in their path; in others they skip from one cluster of trees to another. The result may be a mosaic of color, such as the one that is visible here on a mountainside in Yellowstone National Park. Mosaic burns reflect local variations in moisture and flammability, as well as vagaries of wind and weather.

WILLIAM H. ROMME and DON G. DESPAIN have worked together on a number of research projects. Romme is on the faculty of Fort Lewis College in Durango, Colo., and has spent the past 12 summers in Yellowstone studying vegetation ecology. He is spending the current academic year on sabbatical leave at Oak Ridge National Laboratory in Tennessee, where he is testing models of landscape ecology in collaboration with other ecologists. Romme has a B.A. in chemistry from the University of New Mexico and a Ph.D. in botany from the University of Wyoming. Despain has been a research biologist at Yellowstone National Park for 18 years, which may be the perfect setting in which to study the ecology of fire; when the weather is not too cold, most of his time is spent in the field. Despain has a B.S. in botany from the University of Wyoming and a Ph.D. in plant ecology from the University of Alberta.



EXTENSIVE BURNING occurred in Yellowstone National Park during the summer of 1988. The spread and extent of the fires has been mapped with the help of data provided by satellite imagery and aerial photography. On July 21 the fires were still relatively small (*black*); one month later, on August 21, they were raging across the landscape and the fires' perimeters had expanded farther than expected (*gray*); in just one day more square miles burned than had burned during any decade since 1872. By October 2 a total of 720,000 acres within the park's boundaries had burned (*red*). Although the fires did not spread much after the beginning of October, they did not completely die out until November 13, when heavy snows fell. It should be noted that within each one of the burned areas, there were patches of vegetation that burned lightly or not at all.

increases. That in turn creates air-pressure gradients that generate localized winds. Burning tufts of twigs and conifer needles are then swept into the air by the wind and transported as much as a mile and a half away, where they ignite further fires.

Once a fire ignites, the manner in which it spreads is not always predictable: within any given area there are patches of intensely burned areas intermingled with lightly burned and unburned areas. Such a mosaic pattern reflects not only variations in soil moisture and fuel but also vagaries of wind and weather, all of which complicate the fire-fighting campaign.

Concerned by the rapidly spreading fires, officials invited a number of specialists to the park in early August to estimate the probable course of the fires during the coming weeks. With the help of sophisticated computer models and information gleaned from a number of sources—individuals on the front lines of the fire-fighting effort, the weather bureau, vegetation maps and projections of fire behavior under known weather and fuel conditions—the specialists produced models that were thought to reasonably predict the extent of burn in Yellowstone. But August brought weather unlike any that had been experienced in the northern Rockies over the past century, and the models greatly underestimated the total area likely to burn. Almost no rain fell, temperatures remained high (reaching 90 degrees Fahrenheit or higher on some afternoons) and a series of dry, cold fronts brought prolonged high winds to the region.

Under such conditions moisture levels in dead logs and branches, which normally range from 15 to 20 percent, dropped to only 7 percent; in small dead twigs they dropped to 2 percent. Moisture levels that low are known to support large fires; as a consequence, the computer models of fire behavior were modified to compensate for the unexpected conditions. Indeed, by late August seven major fires in the Yellowstone region were burning. On some days they advanced by as much as 10 or 12 miles and burned tens of thousands of acres. By mid-September, when the air finally began to clear, an aerial survey of the entire Yellowstone area revealed that approximately one million acres had been affected by fire, 720,000 of them in Yellowstone Park itself.

It is often said that the fires of 1988 would not have raged so uncontrollably had they been suppressed from the beginning. Did the fault for their unusual behavior lie with Yellowstone's "let it burn" policy? And if so, why had such a policy been established in the first place? To answer those questions, the issue of fire control must be examined from a historical perspective.

In 1872, when Yellowstone was established as the world's first national park, the goal of management, as explained in the enabling legislation, was the "preservation from injury or despoliation of all timber, mineral deposits, natural curiosities, or wonders within said park, and their retention in their natural condition." The destruction of timber by fire was viewed as antithetical to the park's goals, and suppression became a high priority.

It was not until the 1960's that the value of suppression was widely questioned by ecologists and park officials. At that time ecologists began to view natural disturbances, especially fires, as interesting and significant processes in their own right. One such ecologist was Dale L. Taylor, then at the University of Wyoming, who was interested in the species composition of forests at various stages of recovery from fire. He found that species diversity was greatest in young forests, which had replaced forest burned only a few years before. Taylor concluded that eliminating fires from Yellowstone would impoverish the park's wildlife. He also noted that certain species seem to thrive in the aftermath of a fire; the three-toed woodpecker, for example, is common in post-burn areas, where wood-dwelling insects and dead trees suitable for nesting abound.

The shift in ecological thinking about fires coincided with a landmark statement issued in 1963 by a commission appointed to study wildlife management in the U.S. national park system. The Leopold report, named for the head of the commission, A. Starker Leopold, focused explicitly on the dynamic nature of natural ecosystems. The authors recommended that "the biotic associations within each [national] park be maintained, or where necessary re-created, as nearly as possible in the condition that prevailed when the area was first visited by the white man."

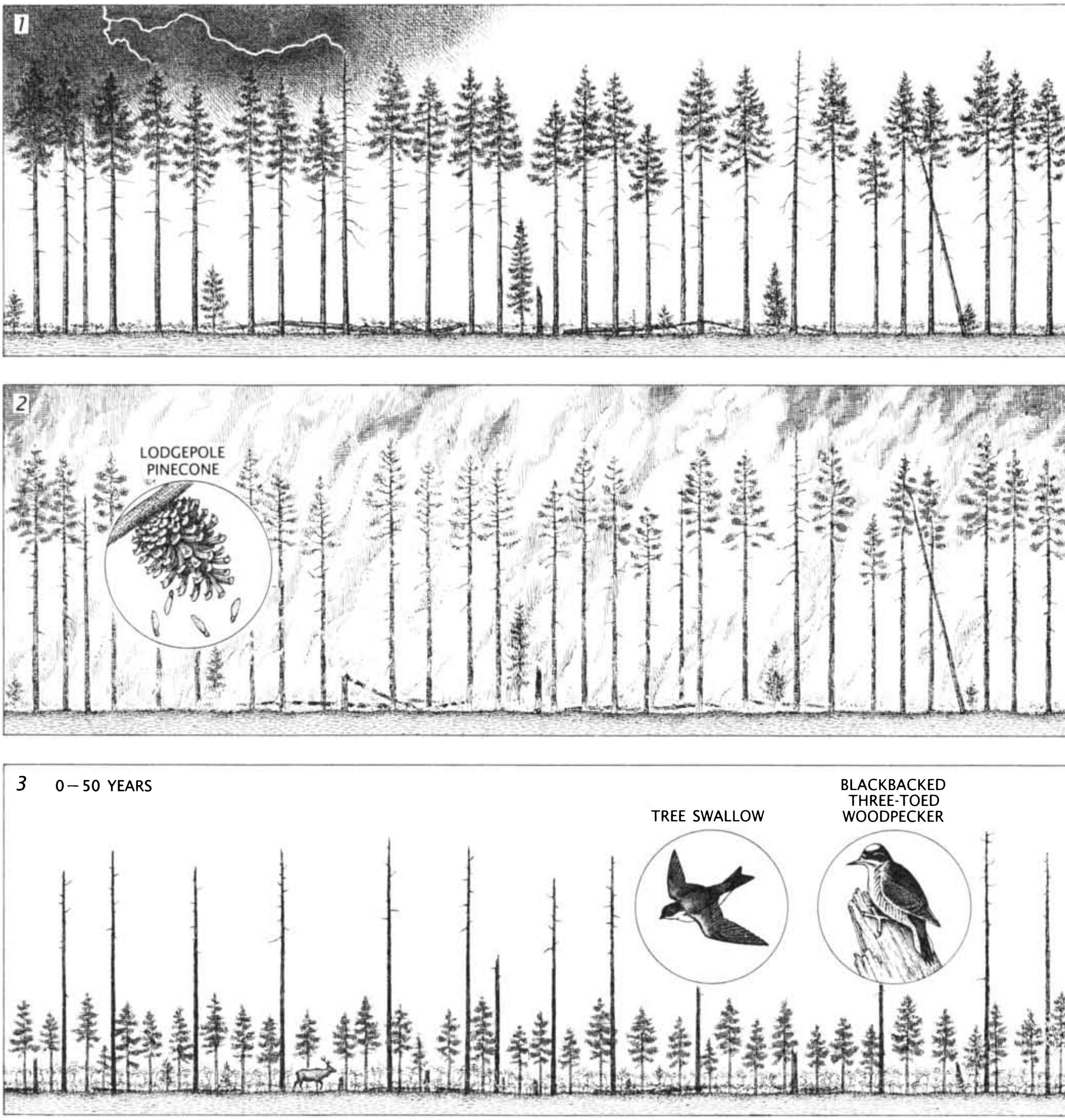
Several kinds of evidence indicated that fire had been important to the Yellowstone ecosystem. Photographs taken during the late 1800's, for example, showed young forests growing amid charred stumps; photographs of the same sites taken 100 years later reveal older, denser growth and little evidence of recent fire. Species such as conifers and sagebrush that are generally susceptible to fire were more abundant in the later photographs, whereas species, such as aspen, that do well in fire-prone areas were less abundant.

In an attempt to retain Yellowstone in a natural condition, park officials decided to initiate an experimental fire program in 1972. Fires ignited by lightning in remote areas would be allowed to burn without interference, although suppression would be practiced elsewhere. In its first few years the program proved so successful that the decision was made in 1976 to extend the program to all of Yellowstone, with the exception of regions inhabited by humans, where fires would still be suppressed. In the nondeveloped areas (about 95 percent of the total acreage of Yellowstone), fires identified as human in origin would be promptly suppressed, as would all lightning fires that might threaten life, property or valuable resources. Lightning fires that posed no apparent threat would be allowed to burn under close surveillance.

Lightning fires that were allowed to burn were designated "prescribed natural fires," indicating that they were within the limits prescribed by the fire-management plan. Fires, regardless of their origin, that had the potential to damage life, property or resources would be declared "wildfires," which meant they were burning outside of prescribed guidelines, and were to be suppressed immediately. The goal of the management directive was to restore fire to the Yellowstone ecosystem yet provide protection for its human inhabitants and resources.

Until the summer of 1988 the natural-fire program was widely viewed as a success. Between 1976 and 1987 some 235 lightning-caused fires were allowed to burn without interference, about 15 a year on average. The vast majority died out after burning no more than 100 acres, and only eight burned more than 1,000 acres. Even the largest fire (which burned 7,400 acres in 1981) was never a threat to lives or resources. Fire was once again a natural process within the Yellowstone ecosystem.

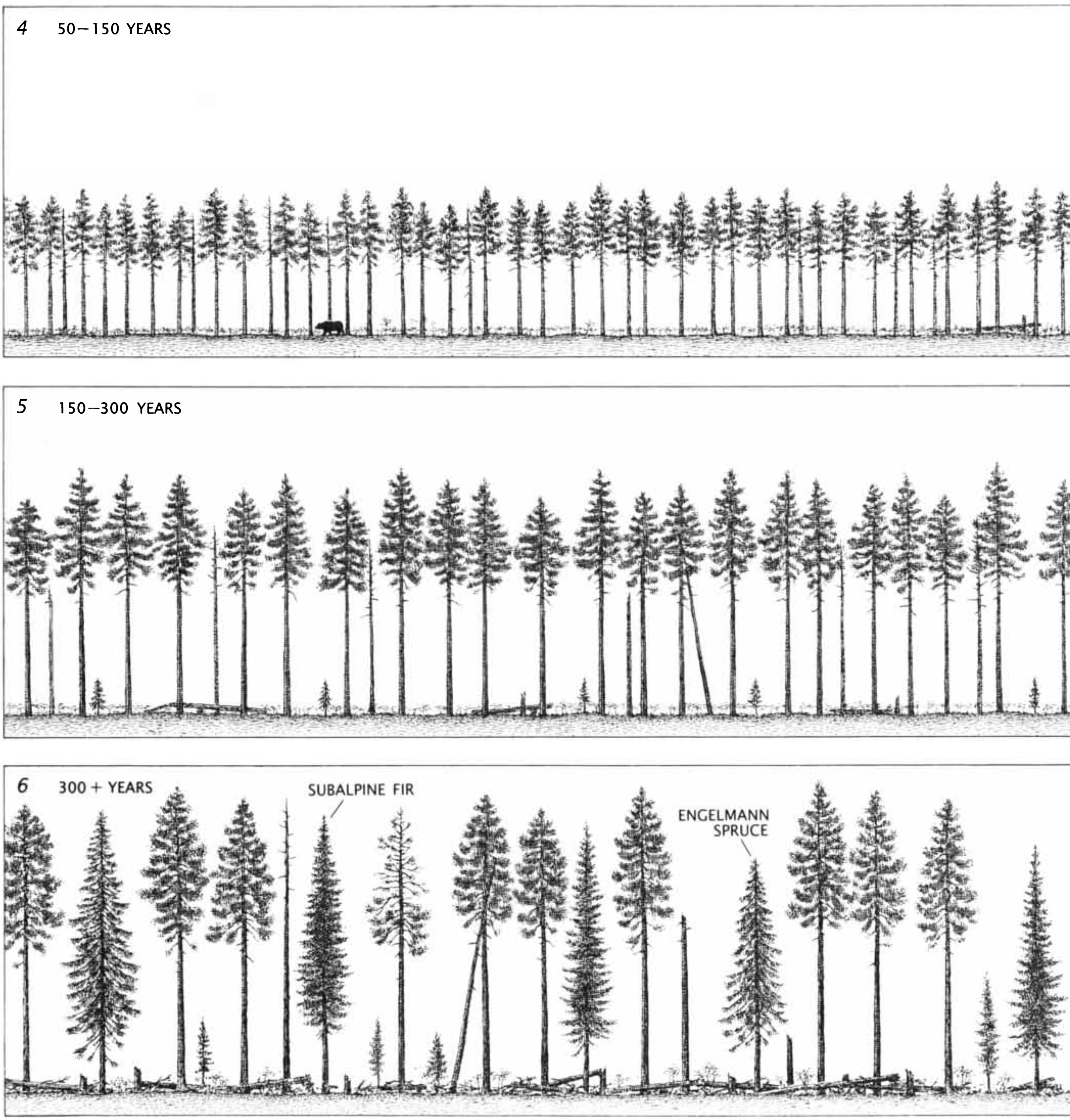
Against the backdrop of such a successful program, the 1988 fires came as a surprise to almost everyone. Were the fires an unusual but nonetheless natural occurrence, or were they the result of careless human behavior and poor management? Some critics have suggested that fire management in the park—if not misguided—was at least poorly designed. Research carried out in other national parks, such as Kings Canyon, Sequoia and Yosemite, indicated that suppressing fires over long periods, say 100 years or more, in ecosystems that would otherwise burn at 10- to 20-year intervals, could lead to the buildup of abnormal amounts of fuel in the form of trees and shrubs in the understory. Such findings suggest that reduction of fuel through prescribed burning be carried out before allowing lightning-caused fires to burn without interference. Was such a fuel-reduction program necessary in Yellowstone? The question, an important one, can be answered—at least in part—by studying the history of fires in Yellowstone.



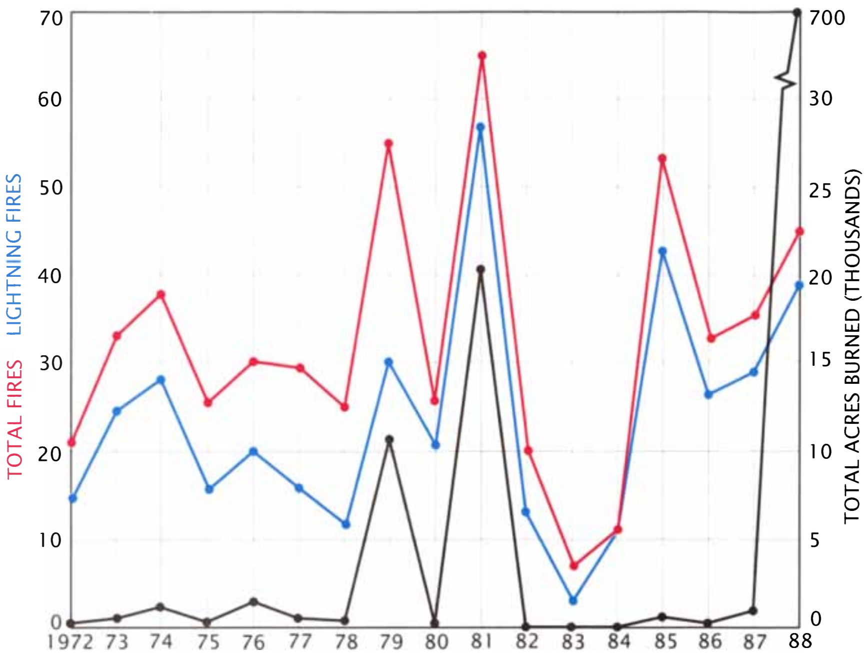
ECOLOGICAL SUCCESSION can be divided into stages, each of which is represented above. Here an old-growth forest in Yellowstone, consisting of lodgepole pine, subalpine fir and Engelmann spruce, has been struck by lightning (*1*). If the ensuing fire is severe (*2*), most vegetation will be destroyed. Lodgepole pinecones protect their seeds during a fire; afterward, they open and release their seeds, which germinate at about the time the roots and seeds of many herbaceous species resprout, thus initiating the first stage of succession (*3*). During that stage the open, sunny forest supports many species, including birds, such as swallows and woodpeckers, and elk. Lodgepole saplings may grow to 15 or 20 feet. After

Douglas B. Houston, an ecologist at Yellowstone National Park, has studied the history of fires in low-elevation areas in Yellowstone. The vegetation in these regions consists primarily of sagebrush grasslands interspersed with forests of Douglas fir and aspen; willows and sedges are typically found along the streambeds. By analyzing fire scars on trees, Houston was able to determine the fire history of the area. Scars form on the trunk where the tree's cambium, or layer of ring-forming cells, has been killed by fire. The wedge-shaped scars run vertically from one to 15 feet up a tree and are quite distinctive.

In regions of the trunk where there is no scarring, annual rings continue to form, and these can be counted to determine the year in which a fire



about 50 years the forest enters the second stage of succession (*4*), which lasts for about 100 years. During this period the pines reach heights from 30 to 50 feet and form dense stands that block the sun. In the third stage (*5*), which starts after some 150 years and lasts for about a century, the lodgepole pines thin out, and second-generation trees such as the Engelmann spruce and subalpine fir appear. Increased sunlight stimulates the growth of vegetation on the forest floor. In the last stage of succession (*6*), when the forest is about 300 years old, the original trees die, and large gaps appear in the canopy. Small trees and dead branches accumulate, and the forest, which is now highly flammable, is once again vulnerable to fire.



NUMBER AND SEVERITY of fires in a region varies from year to year depending on weather conditions (temperature, wind, drought) and a forest's stage of succession. For each year from 1972 to 1988, the total number of fires (*red*), the number of acres burned (*black*) and the number of fires ignited by lightning and designated prescribed natural fires (*blue*) are indicated for Yellowstone National Park. Although the number of fires in 1988 was not unusually high, more acres burned that summer than had burned in any decade since 1872, when such data were first recorded.

occurred. The procedure involves cutting a cross section or wedge from the circumference of the tree and then counting its growth rings under a microscope. Fire scars form only in trees that have been exposed to fire but not killed and thus are somewhat rare; even so Houston was able to find enough scarred trees to put together a history of fire occurrence in his study area. He found that (before the 20th century, when suppression policies were in effect) fires would recur at any one spot about every 20 or 25 years on average, although the interval might range from as little as two to as many as 90 years.

We studied fire history at higher elevations, where the climate is cooler and wetter and where the vegetation is predominantly coniferous. There we found an entirely different situation from the one described by Houston for lower elevations. Our study area consisted of 320,000 acres in the south central section of Yellowstone. By analyzing fire scars, we determined that fires have burned at various locations in this area every decade since at least 1690. Most were small, but extensive fires did occur during a few key periods: from 1690 to 1709, when 19 percent of the study area burned; from 1730 to 1749, when 15 percent of it burned; from 1850 to 1869, when 9 percent of it burned; and finally in 1988, when 26 percent of it burned.

Why, over a time span that lasted almost 300 years—from the early 1700's to the summer of 1988—were there so few large fires? And why, when the fires did appear, were they so spectacular? The answers have to do (at least in part) with the nature of succession.

Young forests in the first stage of ecological succession are dominated by herbaceous plants and young lodgepole pines; dead trees (both standing and fallen) from the pre-burn forest are numerous. Lodgepole pines, *Pinus contorta,* germinate the first year after the fire, reaching five feet in 10 years and from 20 to 30 feet by the time the canopy closes, a sign that the first stage has come to an end. Species diversity is high during this stage, which lasts for about 50 years. The low-lying vegetation remains green and moist throughout the fire season, and the trees are widely separated, which inhibits fire from spreading, and the forest as a whole is not very flammable.

During the next stage the lodgepole pines form dense stands ranging from 30 to 50 feet tall. Natural thinning of the trees takes place as they get bigger, a result of competition for light, water and nutrients. Ground vegetation is sparse and species diversity is low, largely because the trees' upper branches block much of the sun from reaching the forest floor. This stage starts about 50 years after the fire and lasts for about 100 years. There are some rotten logs and dead saplings on the ground, but the treetops rise too far above the forest floor to be easily ignited from below, and so the second stage, like the previous one, is not very flammable.

By the third stage lodgepole pines are less dense, ground vegetation is more abundant and a new generation of pine saplings begins to grow. Other tree species including the subalpine fir, *Abies lasiocarpa,* and the Engelmann spruce, *Picea engelmannii,* may also appear, depending on the habitat and elevation. This stage usually begins some 150 years after the fire and lasts for about 100 years. Vegetation on the forest floor stays green during most of the fire season, so fires tend to spread slowly when they occur. During the later part of this stage, however, an understory of small trees provides a fuel bed from which fires can spread into the canopy.

In the last stage of succession, the original lodgepole pines, which may be 300 years old or more, begin to die, creating an uneven canopy with many gaps. The young saplings, which appeared during the previous stage, grow into the gaps and so change the species composition of the forest. The forest is quite flammable during this stage: young trees proliferate in the understory, woody fuels made up of different species are abundant and the flammable crowns of spruce and fir extend down to the ground, creating conditions under which fire can easily spread to the canopy. Eventually, after two or three centuries, the vegetation in burned regions will again resemble what was present before the fires. In the meantime other areas will burn, and thus a mosaic of forests (with their associated wildlife) in various stages of succession is created.

Theoretically (in the absence of fire) the lodgepole pines that originally colonized the site would die and the forest would consist of just spruce and fir trees; at this point the forest would have reached the final, or climax, stage of succession. Such stands, however, are very rare in Yellowstone; nearly all burn before reaching the climax stage.

One characteristic of Yellowstone forests is that they generally become more flammable as they age. When lightning strikes younger forests, it may ignite fires that smolder on the ground for several weeks, but without an abundant fuel supply the fires will go out before much damage is incurred. In contrast, the accumulation of small trees and dead plant matter in old-growth forests renders them highly susceptible to crown fires. We concluded from our studies that lightning probably has ignited fires in Yellowstone every year but that when young forests cover the landscape, as they did in the 1700's and early 1800's, the fires are usually small and have limited impact.

Our study also showed that by the beginning of the 20th century much of the area that had burned around 1700 had begun to enter the old-growth stage of succession. By 1930 or 1940 the number of acres covered by flammable, old-growth forests was greater than it had been at any time during the preceding 250 years. At that point large fires were almost inevitable. Why, if that was the case, were they so long in coming? It appears that a combination of unusual weather conditions and past suppression strategies is responsible.

AFTERMATH of the 1988 fires reveals a forest of blackened trees, devoid of visible life. The ground is covered by a layer of ashes, which provides fertilizer for the plants that will resprout the following year from seeds and roots below ground.

Weather is a powerful determinant of fire behavior. In a normal summer much of Yellowstone is simply too wet to burn, and most fires, including those in old-growth forests, die out quickly. Even during a dry summer, fires generally do not burn large areas. Although they may smolder into October, they flare up only occasionally. In the dry years of 1976, 1979 and 1981, for example, only 1,600, 11,000 and 20,600 acres burned, respectively. Wind is also an important determinant of fire size. A year such as 1988, with its unique combination of drought and sustained strong winds, probably comes along only every century or so. When weather conditions coincide with a landscape that is covered by large areas of flammable, old-growth forest, the stage is set for extensive burning.

Suppression strategies and their relation to the fires of 1988 are more difficult to evaluate. At lower elevations, and especially along roads and trails, fire suppression may have contributed to an increase in the area covered by conifers and sagebrush and so increased the flammability of the region. The impact of suppression at higher elevations, which constitute 80 percent of the park, is less clear. Before the 1940's, fire-fighting efforts in remote high-elevation regions probably were not very effective. Once a fire was large enough to be spotted,

FLOWERS have sprung up in the wake of the Yellowstone fires. Purple fireweed and other herbaceous species are often the first to appear in a newly burned patch of moist forest. Their presence indicates that the first stage of succession has begun; drier, less fertile sites may lag behind such fertile sites by three to five years.

several hours or a day was often needed for a crew to hike into the area, by which time the fire either had gone out or had grown so large that it could not be extinguished with hand tools. Most likely, many fires went out before they were even detected.

The situation changed considerably after World War II with the advent of more sophisticated fire-detection and fire-fighting technologies—patrolling aircraft, helicopters carrying water and fire fighters, slurry bombers (modified bomber planes that carry large quantities of flame-retarding chemicals) and smoke jumpers (fire fighters who parachute from planes and get to fires soon after they start). Although fire suppression became much more effective, it was practiced only for a limited period, namely from the 1940's until the new management directive went into effect in the 1970's.

Our conclusion is that the fires of 1988 were more or less a natural event in the ecological history of Yellowstone Park, a perturbation such as might occur every 200 or 300 years. Data from our study area indicate that the fires of 1988 in fact probably did not behave very differently from the ones that burned across the same area around 1700.

It seems that unusually dry, hot and windy weather conditions in July and August of 1988 coincided with multiple ignitions in a forest that was at its most flammable stage of succession. Yet it is unlikely that past suppression efforts were a major factor in exacerbating the Yellowstone fires. If fires occur naturally at intervals ranging from 200 to 400 years, then 30 or 40 years of effective suppression is simply not long enough for excessive quantities of fuel to build up. Major attempts at suppression in Yellowstone forests may have merely delayed the inevitable.

Although the fires of 1988 may not be unique from a historical perspective, they do provide biologists with an unprecedented opportunity to study large-scale ecological distubances. Yellowstone National Park, which is managed as a natural area, provides an appropriate setting for assessing both the short- and long-term effects of fire on a biological community.

Large animals such as elk and bison seem to have been affected little during the fires: they simply moved out of the way. Only about 350 elk and nine bison are known to have perished, out of herds numbering more than 30,000 and 2,500, respectively. Those that died were felled mostly by smoke inhalation during a few days in early September when the fires were moving at incredible speeds, 10 miles or more a day. Other animals, such as hawks, were affected favorably: by hovering just beyond the flames, they could catch small animals as they were displaced from their habitats.

GRAZING VERTEBRATES, such as the bison that live in Yellowstone National Park, are among the first animals to invade a forest after it has burned. They come to feed on grasses and other herbs that often grow vigorously in the aftermath of a severe fire.

From a longer-term perspective, animals, like plants, are key players in the game of succession. Insects and other invertebrates will colonize trees killed by fire as soon as the forest cools. Deer mice return at about the same time, followed by birds, many of which nest as readily in burned forests as they do in unburned ones. Woodpeckers come seeking an abundant supply of bark beetles and other tree-dwelling insects; tree swallows come to nest in the cavities excavated by woodpeckers. As the density of forage increases (stimulated by soil nutrients and sunshine), bears and elk arrive to eat the plants and insects that proliferate.

After a fire the response of plants—both short- and long-term—is similar to that of animals. Where burning is light, few plants die, and the community usually returns to its pre-burn condition within from three to five years. Where canopy burn is extensive, all aboveground vegetation is often killed. In many cases, however, reproductive structures, such as roots, rhizomes and seeds, will survive below ground; as a consequence, many herbs and low shrubs grow vigorously following a fire, sometimes within the same season. Trees, on the other hand, with the notable exception of aspen, generally do not resprout and so are killed by a forest fire.

Some herbs and shrubs are actually more vigorous and productive after a fire, when competition for light, water and nutrients is reduced, than they were before. In some situations, undesirable weed species, such as the spotted knapweed and the Canadian thistle, may respond to the reduced competition by establishing themselves in areas where they previously did not grow.

The lodgepole pine is a particularly interesting species. Although it must regrow from seed, some of its pinecones are serotinous—they remain closed at maturity, opening and releasing their seeds only after they have been burned. Fire also creates an ideal habitat for lodgepole-pine seeds, which seem to grow best in open, well-lit spaces, where competition from other plant species is minimal.

In addition, fire has a profound impact on nutrient cycling within a community. Nutrients such as nitrogen, phosphorus and calcium, which are usually tied up in organic matter, are released when the organic matter burns. Instead of being lost from the ecosystem, however, they tend to be

Imagine a computer that lets you subtract hours from business trips, multiply sales calls, or divide a week between meetings across a continent.

Now imagine that it's run by an expert staff and globally networked among over 135 cities worldwide.

You've just imagined a KLM jet—an ingenious device for boosting the productivity of business trips.

KLM, for example, operates one of the most modern fleets of passenger aircraft in the world.

Our delightfully uncrowded hub in Amsterdam has been voted Europe's best airport 7 years running.

And no other airline offers seventy years of experience attending to such routine but crucial details as ticketing, meal service or maintenance.

After all, you wouldn't settle for a computer that wasn't state of the art. Why choose an airline by a lesser standard? Call your travel agent or KLM. The airline of the seasoned traveler.

The Reliable Airline **KLM**
Royal Dutch Airlines

Southwest: 800-392-2032 ext. 303; Midwest: 800-621-8940 ext. 200; West: 800-262-1509 ext. 332; Southeast: 800-451-5556 ext. 319; East: 800-652-5266 ext. 3048 conserved by biological activity. Bacteria, fungi and other microorganisms in the soil increase their rates of nutrient uptake in response to the augmented supply; moreover, warmer temperatures created by the greater heat absorption of the blackened ground and by the removal of shade also serve to increase the metabolic activity of soil organisms and so increase the rate of uptake still further.

A second nutrient-conserving mechanism, currently being studied by Samuel J. McNaughton of Syracuse University and Linda L. Wallace of the University of Oklahoma, is related to the shift that occurs in community composition from predominantly woody plant species to predominantly herbaceous ones. Herbaceous plants tend to absorb nutrients more rapidly and recycle them faster than do woody plants. Consequently, the grasses that dominate a burned area accumulate nutrients rapidly. Nutrient cycling is further accelerated by elk and bison, who eat the grasses, returning them to the soil as waste.

Nutrients that are lost through runoff may be intercepted by riparian, or streamside, vegetation such as willows or sedges. Riparian areas often escape burning because they are damp, but even when the vegetation in those areas does burn, it tends to regenerate quickly. According to G. Wayne Minshall of Idaho State University, the flow of nutrients into a stream after a fire, combined with an increase in water temperature and sunlight, leads to an overall increase in the metabolic activity of aquatic organisms living in the stream and so boosts the stream's productivity. Over the long run such periodic influxes of nutrients may also help sustain the algae that form an important base of the food chain in large lakes. An increase in algae will support greater numbers of zooplankton, which in turn will support trout, on which pelicans, ospreys, grizzly bears and other predatory animals feed.

In reflecting on the Yellowstone fires of 1988, one can ask what changes, if any, should be made in existing fire-control programs. Two independent teams of scientists and managers were appointed in the fall of 1988 to reevaluate suppression and other aspects of fire management in Yellowstone. The Fire Management Policy Review Team, appointed by the secretaries of agriculture and the interior, was composed of 10 people representing federal land-management agencies and the Association of State Foresters. Their final report, the Philpot-Leonard report (after Charles Philpot and Brad Leonard, who were cochairmen of the team), was released on December 20, 1988. The second team, appointed by the National Park Service, was called the Greater Yellowstone Area Post-Fire Ecological Assessment Panel. It was composed of 13 people, mostly academic scientists with experience in ecology and watershed research. Their conclusions were presented in the Christensen report, named for Norman L. Christensen of Duke University, which was submitted to the National Park Service on April 15, 1989.

Both groups concluded that some kind of natural-fire program, in which lightning-caused fires are allowed to burn under certain conditions, is appropriate and necessary for maintaining the wilderness value of parks and other refuges. But the Christensen report pointed out that in contrast to the past, when fires ranged freely over North America, the remaining tracts of wilderness are now small, and so nature cannot be allowed to run its course with total freedom. Even though large fires may have occurred in the past, that fact alone does not necessarily make them acceptable today.

The Philpot-Leonard report also emphasized that the risks posed by fires to communities outside the parks should be more clearly recognized. The 1988 fires caused hardship for park visitors and local residents alike. Many tourists were unable to visit Yellowstone as they had planned, and those who went found roads closed and views obscured by smoke. Some of the businesses in the "gateway" communities—the towns on the borders of the park, where tourism is an economic cornerstone—suffered from a drop in visitors. Smoke not only affected visibility in the area but also posed health problems for residents in nearby communities, where particulate levels exceeded the standards of the Clean Air Act.

Should all lightning fires be suppressed and replaced with prescribed fires ignited at predetermined times and places? Both groups recommended that the feasibility of setting fires in developed areas (to reduce fuels and thus the risk of fire) be studied. They did not recommend that natural fires in wilderness areas be replaced with prescribed manager-ignited fires. To begin with, the technology and expertise to carry out such programs safely are not adequately developed: high-intensity canopy fires can easily burn out of control, and low-intensity fires would not simulate the ecosystem's historically important fires. Moreover, the ecological role of fire is not understood well enough to be reproduced accurately. In particular, the issue of fire size is under debate. Are many small fires, for example, the equivalent of one large fire? That question is one that ecologists have only begun to address seriously.

In devising management strategies, it is important to note that as the first national park and one of the largest in the world, Yellowstone is always in the public eye. Witness the press coverage during the summer of 1988. Although 720,000 acres burned in Yellowstone, more than two million acres burned elsewhere in the continental U.S., and 2.2 million acres burned in Alaska. Yet Yellowstone captured most of the headlines. Perhaps the emphasis on Yellowstone was fortuitous, for the problems and issues that beset Yellowstone are representative of the threats and challenges facing all national parks and wilderness areas. Nonetheless, sensational media reports that reinforce erroneous attitudes instead of clarifying important issues do not serve the public well.

Perhaps the most important lesson to come from the summer of 1988 was a reaffirmation of the obvious but often overlooked fact that nature is complex, subtle and not always easily controlled. One of the irreplaceable virtues of wilderness areas is that they provide a setting in which the fundamental workings of nature can be understood. Continuing research in the aftermath of the Yellowstone fires will provide a wealth of understanding about the ecological processes responsible for the diversity and integrity of this priceless area.

FURTHER READING

WILDFIRES IN NORTHERN YELLOWSTONE NATIONAL PARK. Douglas B. Houston in *Ecology,* Vol. 54, No. 5, pages 1111–1117; Late Summer, 1973.

SOME ECOLOGICAL IMPLICATIONS OF FOREST FIRE CONTROL IN YELLOWSTONE NATIONAL PARK, WYOMING. Dale L. Taylor in *Ecology,* Vol. 54, No. 6, pages 1394–1396; Autumn, 1973.

FIRE HISTORY OF A SEQUOIA-MIXED CONIFER FOREST. Bruce M. Kilgore and Dan Taylor in *Ecology,* Vol. 60, No. 1, pages 129–142; February, 1979.

LANDSCAPE DIVERSITY: THE CONCEPT APPLIED TO YELLOWSTONE PARK. William H. Romme and Dennis H. Knight in *BioScience,* Vol. 32, No. 8, pages 664–670; September, 1982.

YELLOWSTONE VEGETATION. Don G. Despain. Roberts Rinehart, Inc., Publishers, in press.

You know his name. And, you're probably familiar with his work. But who *was* Richard Feynman? What made him one of the most extraordinary physicists of his day? Find out in *The Essential Feynman*. Here, in a superb four-volume collection, you'll find the complete man revealed. Best of all, for a limited time only, you can get this unique chronicle of Feynman's life, theories, and accomplishments for the special, low price of just $2.95.

QED
The Strange Theory of Light and Matter

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

Publisher's Price: $18.50

The Character of Physical Law

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

Publisher's Price: $17.50

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

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

Publisher's Price: $15.95

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

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

Publisher's Price: $17.95

MEMBERSHIP BENEFITS • In addition to getting *The Essential Feynman Collection* for only $2.95 when you join, you keep saving substantially on the books you buy. • Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of 65% off the 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.

The Library of Science
Riverside, New Jersey 08075

Please accept my application for trial membership and send me *The Essential Feynman Collection* (00954) billing me only $2.95, plus shipping and handling. I agree to purchase at least three additional Selections or Alternates over the next 12 months. Savings range up to 30% and occasionally even more. My membership is cancelable any time after I buy these three additional books. A shipping and handling charge is added to all shipments.

No-Risk Guarantee: If I am not satisfied—for any reason—I may return *The Essential Feynman Collection* within 10 days. My membership will be canceled, and I will owe nothing.

Name ______________________________

Address ______________________ Apt. ____

City ______________________________

State ______________________ Zip ______

(Books purchased for professional purposes may be a tax-deductible expense. Offer good in Continental U.S. and Canada only. Prices slightly higher in Canada.)

Scientific American 11/89 2-CL9

Double-Beta Decay

The future of fundamental theories that account for everything from the building blocks of the atom to the architecture of the cosmos hinges on studies of this rarest of all observed radioactive events

by Michael K. Moe and Simon Peter Rosen

Nature teases us with its most intimate secrets. Tossing out a clue to catch our interest, it then strews our path with obstacles and false leads. Having followed one such tortuous trail for five decades, experimentalists recently succeeded in detecting the fingerprint of double-beta decay—the rarest radioactive event ever observed in the laboratory. The study of double-beta decay bears on both the fate and fabric of the universe.

In a double-beta event, two neutrons decay simultaneously into two protons, two beta rays (electrons) and two antineutrinos (antimatter versions of the elusive particles called neutrinos). Experimentalists are now searching for another form of double-beta decay, one that does not produce neutrinos or antineutrinos. If such an event is found, it could unravel one of nature's great mysteries: What, if anything, is the mass of the neutrino?

The Standard Model of the elementary particles and the basic forces suggests that this electrically neutral particle should have no mass and should always accompany each electron in double-beta decay. But the model is incomplete at best. Although it successfully accounts for interactions that arise from two of nature's four fundamental forces (the electromagnetic and the weak forces), it does not incorporate the remaining two (the strong force and gravity).

Many theories that go beyond the Standard Model posit that the neutrino should have a definite mass. Double-beta events that take place without releasing any neutrinos or antineutrinos would be a direct consequence of neutrino mass. Although the mass these theories assign to the neutrino is at least 10,000 times less than the mass of an electron, neutrinos pervade the cosmos. If they have mass, they could be a major component of the mysterious dark matter that affects the evolution of galaxies and perhaps the universe as a whole through its gravitational pull [see "Dark Matter in the Universe," by Lawrence M. Krauss; SCIENTIFIC AMERICAN, December, 1986].

Radioactivity in general results from the instability of atomic nuclei. If changing a neutron into a proton in the nucleus will result in a lighter atom, then the transformation takes place by single-beta decay, which releases a single electron and a single antineutrino. The difference in mass between the parent atom and its daughter is the energy available to the ejected electron and antineutrino, according to Einstein's equivalence principle of mass and energy. Conversely, if the neutron-to-proton change would result in a more massive atom, then energy conservation will not allow beta decay.

The new proton created by the beta decay alters the chemical properties of the atom and changes it to the next higher element in the periodic table. The negative charge of the ejected electron balances the positive charge of the created proton, thereby fulfilling the law of conservation of charge. (In a rare reciprocal process, a proton is transformed into a neutron, the atom moves one lower in the periodic table and a positive electron—a positron—is emitted, accompanied by a neutrino.)

The observed form of double-beta decay has the same end result as a sequence of two single-beta decays, but it cannot occur as two separate decays because the first beta decay is energetically forbidden: it would create a daughter nucleus that is heavier than the parent. Only the product of the second decay is lighter than the original nucleus. The two single-beta decays must occur simultaneously by a process of quantum-mechanical tunneling through the energy barrier presented by the first decay. The decay products of the first step—a virtual one—do not materialize until the second step is completed. Double-beta decay always releases two beta rays: fast electrons or positrons. Throughout the history of physics, however, it has never been clear whether neutrinos must always accompany the beta rays.

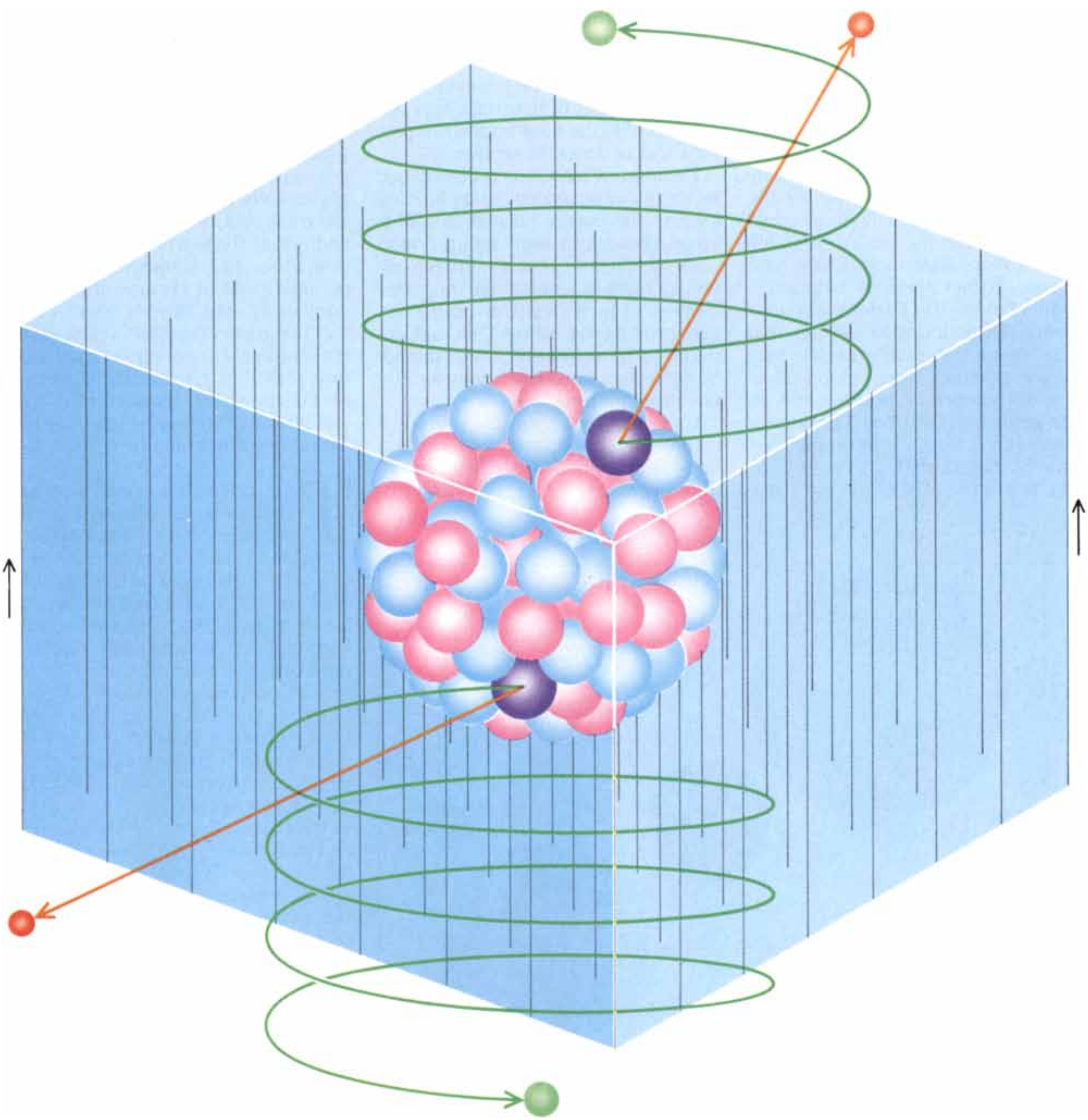
The original clue to the neutrino's existence was the observation of missing energy in single-beta decay. If the mass difference between parent and daughter atoms had been converted into energy for the beta ray, then the beta-ray energies from decaying nuclei of the same type should all have been identical. Yet the observed electrons all had different energies, always less than the energy equivalent of the mass difference. Wolfgang Pauli suspected in 1930 that nature was not actually breaking the law of energy conservation but was concealing an unknown particle. Later dubbed the neutrino, this particle was carrying away some of the energy in beta decay. The neutrino was not easy to detect, however, because it interacted only

MICHAEL K. MOE and SIMON PETER ROSEN have approached double-beta decay from opposite directions. Moe is an experimentalist who holds a B.S. from Stanford University and a Ph.D. from the Case Institute of Technology. Now at the University of California, Irvine, he has been searching since 1971 for double-beta events. In 1987 Moe and two colleagues were the first to observe a double-beta event directly. Rosen has tackled beta decay from the theoretical side. He got his B.S. and Ph.D. from the University of Oxford. In 1959 he and the late Henry Primakoff published a paper that set the stage for the ensuing investigations of double-beta decay. Now at the Los Alamos National Laboratory, Rosen's interests have extended to include the forces behind double decay, namely, weak interactions.

weakly with matter—so weakly, in fact, that it was not observed until 1956 when Clyde L. Cowan, Jr., and Frederick Reines of the Los Alamos Science Laboratory captured a few neutrinos emanating from a nuclear reactor.

According to the Standard Model, the neutrino accompanying a negative beta ray is the distinct antiparticle of the one accompanying a positive beta ray (just as the positron is the distinct antiparticle of the electron). Theories that go beyond the Standard Model and assign a mass to the neutrino, however, predict that the particle emitted with a negative beta ray should be the same as the one emitted with a positive beta ray. In other words, the neutrino would be its own antiparticle. How can we tell whether these predictions are right?

Double-beta decay is the ideal process in which to seek an answer to this question. If the neutrino has mass and is its own antiparticle, then the neutrino emitted in the first stage of the process might be reabsorbed in the



NUCLEAR SIGNATURE of double-beta decay emanates from the nucleus of an atom of selenium consisting of 48 neutrons (*blue and purple*) and 34 protons (*red*). Two of the neutrons (*purple*) decay simultaneously into two protons and in the process generate two beta rays (electrons) (*green*) and two antineutrinos (*orange*). An external magnetic field (*gray*) causes the paths of the ejected electrons to spiral. The double spiral is an observable signal of the double-beta event. The resulting atom has two more protons and two fewer neutrons than the original state; it has been transformed from selenium to krypton.

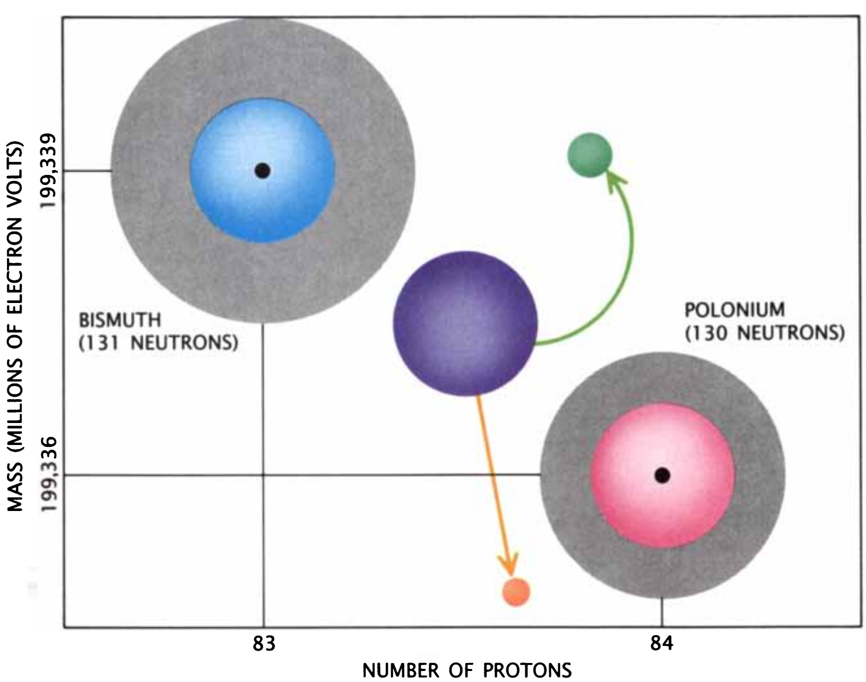
second, yielding a form of decay in which no neutrinos materialize. In this case one expects that in at least some double-beta decays the two electrons will be emitted alone, unaccompanied by any neutrinos. If, on the other hand, the neutrino and antineutrino lack mass and are distinct objects, then the neutrino emitted in the first stage of double-beta decay can never be reabsorbed in the second stage, and two neutrinos must always accompany the two electrons. A neutrino that is its own antiparticle is called a Majorana neutrino after the physicist Ettore Majorana, who first thought of the idea.

Double-beta decay in any form once lay in the realm of theory. Nature tossed out its first clue to the phenomenon in the 1930's. Some candidates for beta decay could still be found in the earth's crust, billions of years after they should have become extinct through radioactive disintegration. The German physicist Werner Heisenberg pointed out that these seemingly stable nuclei were "even-even": they were composed of an even number of protons and an even number of neutrons. The nuclear strong force tends to draw together pairs of like particles more tightly than it binds pairs of unlike particles. Because every nucleon (protons and neutrons) in an even-even nucleus can pair with another nucleon of the same type, such nuclei are more stable than neighboring "odd-odd" nuclei that contain the same total number of nucleons but consist of odd numbers of neutrons and protons. For instance, the 34 protons and 48 neutrons of selenium 82 are more tightly bound than the 35 protons and 47 neutrons of bromine 82. (Elements are distinguished by their chemical properties, which are related to the number of protons. Isotopes of an element have the same number of protons but a varying number of neutrons; they are identified by the total number of protons and neutrons in an atom.)

The mass of a nucleus is diminished by the mass equivalent to its binding energy: the energy needed to pull it apart. Therefore, a more tightly bound nucleus is lighter than a more loosely bound nucleus containing the same number of nucleons. Thus, selenium, being the lighter element, should not decay to form bromine, even though bromine has one fewer neutron and one more proton.

This stability argument was not a complete solution, however. It can and does happen that a particular even-even nucleus is lighter than its nearest odd-odd relative (one more proton, one fewer neutron) but heavier than the next nearest even-even relative (two more protons, two fewer neutrons). Although selenium 82 is lighter than bromine 82, it is heavier than krypton 82 (36 protons and 46 neutrons). Therefore, although beta decay from selenium to bromine is forbidden energetically, double-beta decay from selenium to krypton should indeed take place.

In 1935 Maria G. Mayer of Johns Hopkins University, at the suggestion of her colleague Eugene P. Wigner of Princeton University, calculated the half-life for double-beta decay—the time required for half of the atoms of a particular isotope to decay. In Mayer's scheme, two neutrons decay simultaneously into two protons, two electrons and two antineutrinos. The addition of the two protons places the new atom two elements up in the periodic table (it changes the atomic number by two). Mayer's result was a half-life greater than 10^{17} years, which is exceedingly slow even on a geologic time scale. Thus, the rarity of double-beta decay would account for the original puzzle of why some unstable isotopes were still around.



SINGLE-BETA DECAY can transform the isotope bismuth 214 into polonium 214. The transformation occurs when a neutron decays into a proton, emitting in the process an electron and an antineutrino. The isotope number—the number of protons added to the number of neutrons—remains the same, but the loss of mass (energy) in the decay process becomes the energy available to the ejected electron and antineutrino.

That calculation alone was not enough to establish the reality of double-beta decay. Mayer's colleagues began to contemplate ways to confirm the existence of double-beta decay. The predicted rate of double-beta decay was far too slow to tempt any experimentalist into trying to observe it directly in the laboratory. The experimentalists recognized, however, that there might be some hope of finding indirect evidence for double-beta decay by looking for telltale accumulations of daughter products in geologically old minerals rich in the parent. Normally the infinitesimal buildup of daughter atoms would be swamped by hordes of identical atoms present in the minerals from the time of their formation. Atoms of noble gases, however, are largely excluded from minerals on crystallization because of their volatility and chemical inertness; any such atom produced within the sample would add to the low, natural abundance of the gas. Fortunately among the isotopes that are candidates for double-beta decay, there are some whose daughters are indeed noble gases.

Double-beta decay of selenium 82 would yield a noble gas, krypton; the decay of tellurium 128 and tellurium 130, two other double-beta-decay candidates, would produce another noble gas, xenon. Minerals that contain selenium or tellurium should therefore accumulate krypton or xenon over time. To be sure, the amount of the gas produced in the billion-year history of a mineral would be small—less than one part in 100 million, if Mayer's estimates were essentially correct.

In 1949 Michael G. Inghram and John H. Reynolds, both of the University of Chicago, pioneered a technique for examining fossil gases trapped within ancient selenium and tellurium ores. They released the gases into a mass spectrometer to determine their composition. In 1968, after some refinement of this geochemical method, Till Kirsten, now at the Max Planck Institute in Heidelburg, the late Oliver A. Schaeffer of the State University of New York at Stony Brook and Elinor F. Norton and Raymond W. Stoenner, both of the Brookhaven National Laboratory, found a definitive excess of xenon 130 in 1.3-billion-year-old tellurium ore. This result provided the earliest undisputed evidence that double-beta decay actually occurs.

From the age of the ore and the fraction of tellurium that had decayed to xenon, the half-life of double-beta decay was determined for tellurium 130. There were two important problems with these geochemical experiments. For one, processes other than double-beta decay might have created small amounts of xenon. For another, small amounts of the gas might have been lost from the ore through slow diffusion processes or sudden catastrophic events that heated the ore. The investigators were able to argue that these problems were not serious, but doubts persisted.

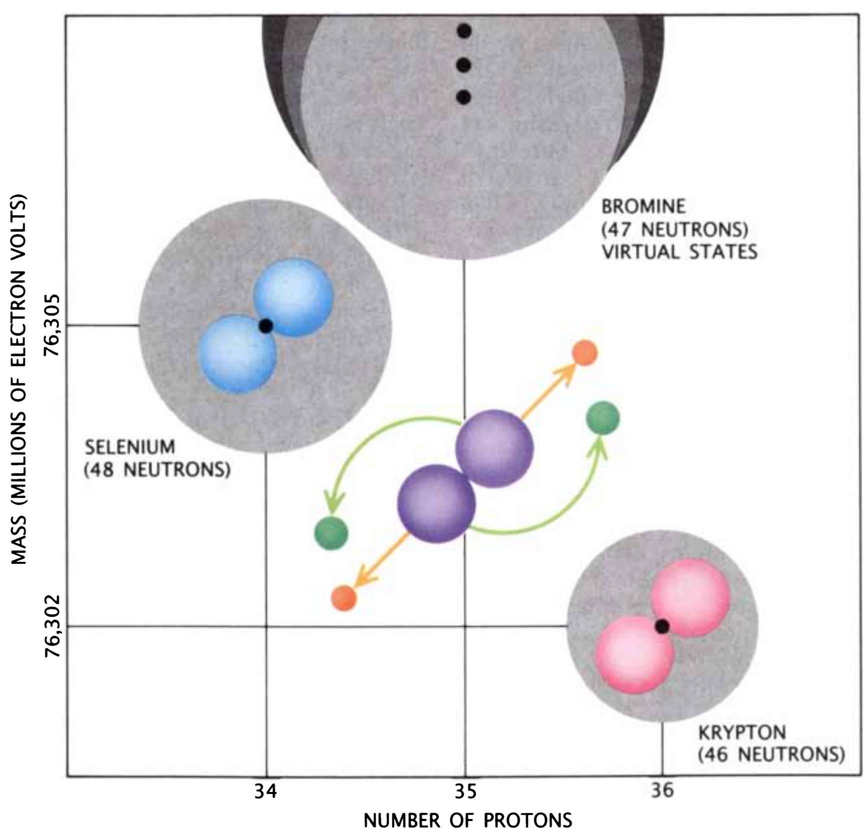
In 1939, four years after Mayer's calculations of theoretical half-lives were published, Wendell H. Furry of Harvard University suggested the possibility that double-beta decay could take place without the emission of neutrinos. Although conservation of energy and momentum required the emission of a neutrino in single-beta decay, there was no corresponding requirement for neutrinos in double-beta decay. Energy and momentum could be conserved in a decay releasing two electrons only. Furry recognized that if neutrinos were Majorana particles—identical to their antiparticles—then no-neutrino double-beta decay could compete with Mayer's two-neutrino double-beta decay. Furry estimated that no-neutrino double-beta decay should occur a million times more frequently than the two-neutrino mode. Nevertheless, the half-life was still on the order of 100 billion years; double-beta decay would still be rare enough to account for the apparent stability of even-even nuclei.

It seemed that the large difference between the predicted lifetimes of the two decay modes might make it possible to determine whether no-neutrino decay actually was taking place. Edward L. Fireman of Princeton University took up the challenge in 1948. He obtained two samples of tin, one artificially enriched in the double-beta-decay candidate tin 124 and the other depleted in that isotope. He placed each sample between a pair of Geiger-Müller tubes (Geiger counters) so each tube could receive one of the two electrons from double-beta decay; consequently, the tubes would fire simultaneously whenever a double-beta decay occurred. He found that simultaneous firing of the tubes took place significantly more often with the enriched sample than with the depleted one. From the data he calculated a half-life much closer to Furry's value than to Mayer's. He concluded that he had observed the no-neutrino mode.

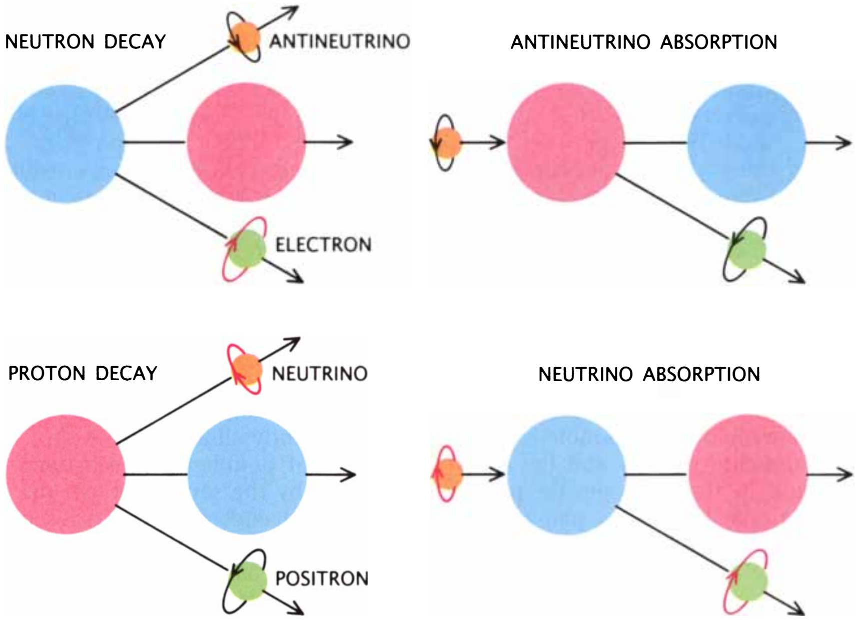
The excitement following this result was short-lived, however. Experiments performed a few years later, including one by Fireman himself, were unable to confirm that the result was actually caused by double-beta decay. Fireman finally conceded that his original results were probably distorted by a small trace of a radioactive impurity in the enriched sample of tin 124.

Until recently all attempts at direct detection of double-beta decay were frustrated by the same problem that Fireman had encountered. Traces of radioactive elements in quantities as small as one part per billion easily masked significant events in double-beta-decay sources, because the half-life of the decay—even by optimistic estimates—was at least a billion times longer than that of common radioactive decays.

The stumbling blocks nature placed



DOUBLE-BETA DECAY from selenium 82 to krypton 82 occurs by way of bromine 82. As in all double-beta decays, the parent isotope is heavier than the final product, but both are lighter (less energetic) than the intermediate state. Single-beta decay of selenium to bromine is energetically forbidden, but double-beta decay to krypton, through a "virtual" intermediate state in bromine, is allowed by the uncertainty principle. The final state of krypton appears after two neutrons become two protons, with the emission of two electrons and usually (perhaps always) two antineutrinos.



NEUTRINOS are emitted and absorbed only in certain ways, according to experiments and the Standard Model. When a neutron (*top left*) inside the nucleus transforms into a proton, it emits an electron and a right-handed antineutrino. (A right-handed particle spins clockwise, as seen from behind.) When a proton (*bottom left*) transforms into a neutron, it emits a positron and a left-handed neutrino. A proton (*top right*) can absorb only a right-handed antineutrino (whereupon it transforms into a neutron). A neutron (*bottom right*) can absorb only a left-handed neutrino.

in the way of experimentalists would soon be overshadowed by a more serious problem. In 1957 Chien-Shiung Wu of Columbia University and her colleagues at the National Bureau of Standards conducted an experiment based on the theoretical work of Chen Ning Yang of the Institute for Advanced Study in Princeton, N.J., and Tsung-Dao Lee of Columbia University. Its results later cast doubt on the very existence of the no-neutrino mode. This experiment and others soon revealed that all antineutrinos emerge from single-beta decay with the same "handedness" [see "A Flaw in a Universal Mirror," by Robert K. Adair; SCIENTIFIC AMERICAN, February, 1988].

The implications of the discovery for physics and for double-beta decay arose from basic properties of certain fundamental particles. Neutrinos, electrons, protons and neutrons all possess an intrinsic spin. One can think of them as miniature tops spinning around some internal axis as they travel through space and inside atoms. The top has two possible orientations. As the top travels away from an observer, it spins either clockwise or counterclockwise. To put it more precisely, in the clockwise, or right-handed, arrangement the vector that describes the spin is aligned in the same direction as the vector that describes the movement or momentum. In the left-handed arrangement, the spin vector points in the direction opposite to the momentum vector. Therefore, particles such as electrons and neutrinos can be either right- or left-handed.

In some interactions, a right-handed particle is just as likely to be generated as a left-handed one: what spins clockwise can equally well spin counterclockwise. When the probability of one outcome is the same as the probability of its mirror image, parity is said to be conserved.

Wu discovered that in single-beta decay parity is not conserved. Her work led to other demonstrations that the antineutrino that accompanies a negative beta ray (an electron) is always right-handed. Furthermore, the neutrino that accompanies a positive beta ray (a positron) is always left-handed. If parity were conserved in beta decay, then the occurrence of right- and left-handed configurations would be equally probable for each kind of decay.

The discovery of parity nonconservation in beta decay implied that no-neutrino double-beta decay would be a very unlikely process. Theorists arrived at this conclusion by considering models of the no-neutrino mode of double-beta decay. No-neutrino decay releases two electrons only. (The release of two positrons is also possible but is expected to be much rarer.) In the first half of the no-neutrino process, one neutron releases an electron and a right-handed antineutrino, which must then be absorbed within the nucleus [*see illustration at left*]. All evidence suggests that a right-handed antineutrino can be absorbed only by a proton. When this happens, the proton ejects a positron and thereby becomes a neutron. This process would not lead to double-beta decay, because the product has the same nucleus as the original state and hence has the same mass. If the right-handed antineutrino could become a left-handed neutrino, however, a neutron could absorb it. The neutron would then decay into a proton, ejecting a second electron, unaccompanied by a neutrino. An observer would see two neutrons decaying into two protons, ejecting in the process two electrons and no neutrinos.

Furry's theory provided a mechanism for the antineutrino-to-neutrino conversion because it assumed that neutrinos are Majorana particles. The distinction normally made between neutrino and antineutrino may be an artificial one based only on the different handedness with which they emerge from different processes. Nevertheless, when the consistent handedness was considered, the no-neutrino mode appeared to be forbidden, regardless of whether the neutrino was a Majorana particle or not! At this point, most experiments on double-beta decay were abandoned.

With theoretical developments of the late 1970's, support for the idea that neutrinos have mass began to emerge. Massive neutrinos would remove the handedness impasse that had theoretically prevented no-neutrino double-beta decay. The rate of decay allowed by neutrino mass would be much slower than Furry's original prediction, but decay could proceed at some very low level determined by the size of the neutrino mass. Thus, the search for neutrinoless double-beta decay became a test for neutrino mass as well as for Majorana neutrinos, and a great resurgence of experimental and theoretical activity began.

The relation between neutrino mass and neutrino handedness arises from the special theory of relativity. The theory states that a massless particle always travels at the speed of light. In other words, no frame of reference can be found in which the particle is at rest. As a consequence, a particle that does not have mass cannot be brought to rest and then be set in motion in the

Announcing Grand Prix Sports Sedan. The Four-Door Version Of A Performance And Styling Legend.

This new sports sedan has the road car credentials that have made the Grand Prix™ name renowned on every highway in America for nearly 30 years. As evidence, we submit a 16-valve 2.3L Quad 4 engine, or optional 3.1L V6 with 4-speed automatic. There are also 4-wheel power disc brakes with an optional Anti-Lock Brake System. Other options include sunroof, cast aluminum Sport wheels and a Sport Appearance Package with special grille, fog lamps, red-accented body side moldings and more. Inside, there's a comfortably contoured 40/60 split front seat and a roomy bench in back. Analog instruments are the focal point of Grand Prix Sedan's ergonomic cockpit, and a performance-contoured steering wheel is the touchstone for driving Excitement in the Pontiac tradition. As if the Grand Prix marque wasn't proof enough.

PONTIAC
WE BUILD
Excitement

Call Toll-Free 1-800-888-2997 For More Product Information.

BUMPER TO BUMPER PLUS
3 YEAR/50,000 MILE WARRANTY

opposite direction in order to reverse its handedness.

In contrast, a particle with a nonzero mass never travels at precisely the speed of light. Such a particle can be brought to rest and then have its direction of motion reversed without changing its spin direction. Therefore, the right-to-left transformation required for no-neutrino double-beta decay could take place if neutrinos have mass.

The need for a right-to-left transformation could be circumvented if all cases of beta decay do not conform to the handedness rules that seem to apply to the emission and absorption of neutrinos. Experiments reveal the spin and momentum vectors only to a certain degree of accuracy. It is therefore possible that the decay of a neutron can on rare occasions produce a left-handed Majorana neutrino instead of the usual right-handed one. Furthermore, a neutron may sometimes absorb a right-handed Majorana neutrino instead of the usual left-handed one. Yet these violations of the rules for neutrino emission and absorption would also require the neutrino to have mass. It appears, then, that neutrino mass is required for no-neutrino double-beta decay to occur and, conversely, that the observation of the phenomenon implies that the neutrino must have mass.

Detection of the no-neutrino mode has eluded experimenters, but progress on the formidable problem of distinguishing double-beta decay from background events has led to the direct detection of the two-neutrino mode. In 1987 Steve R. Elliott, Alan A. Hahn and one of us (Moe) of the University of California at Irvine reported these first results, obtained from a sample of selenium 82. This element was chosen because its decay energy is greater than most and geochemical results indicated a relatively short half-life. Because natural sources of selenium contain only 9 percent of the isotope 82, a sample enriched to 97 percent selenium 82 was obtained to provide more source material for double-beta decay and to reduce the risk of contamination.

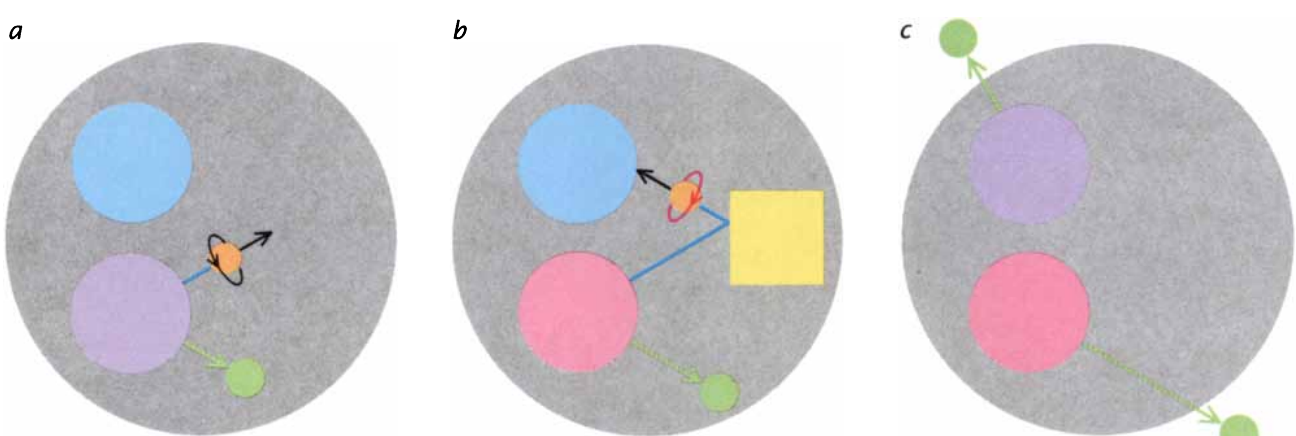
The sample was placed in a device called a time projection chamber [*see upper illustration on page 54B*]. The chamber was filled with gas and immersed in a strong magnetic field. Beta particles traveling through the chamber ionized the gas, leaving tracks that were recorded by sensors. The magnetic field forced each electron to curve in a way that revealed its direction of motion. Any double-beta event would leave a unique signature in the chamber: two electrons spiraling away from the selenium sample.

The uniqueness of the double-beta-decay signature helped tremendously in eliminating extraneous signals. Single electrons released from the selenium could not masquerade as a double-beta event, because they produced only one spiral. An electron and a positron—a pair generated when gamma rays hit the sample—created a double spiral, but the positron curved the wrong way. Two electrons ejected simultaneously from spatially separate events in the source were also easily identified. In order to fool the sensors, two electrons had to emerge from the same point in the sample.

Yet electrons from one insidious process did exactly that. In single-beta decay, the nucleus of the daughter product is often left in an excited state. The extra energy is usually emitted in the form of gamma rays, but sometimes it is transferred to one of the electrons orbiting the atom, especially in heavy atoms such as the decay products of uranium and thorium. Thus, two electrons pop out of a single atom: the beta particle and the ejected orbital electron. The pair imitates double-beta decay almost to perfection.

Fortunately, most uranium and thorium products continue to decay into other elements. After a time they emit other particles, betraying the double-beta-decay imposter. In the case of bismuth 214, a trace contaminant that generates excellent double-beta imitations, an observer does not have to wait very long. Within a millisecond, the daughter product of bismuth 214 emits an alpha particle, composed of two protons and two neutrons, which flags the false event. Alpha particles, which make dense, straight tracks in the gas of the time projection chamber, have little penetrating power—a piece of paper stops them cold. Therefore, the selenium sample for the time projection chamber was made thin to allow the alpha particles to escape and betray the false events. This requirement, unfortunately, limited the mass of the sample and hence the sensitivity of the experiment.

The entire apparatus was enclosed in thick walls of lead, which shielded the chamber from gamma rays that emanate naturally from the laboratory's concrete walls. Because the lead

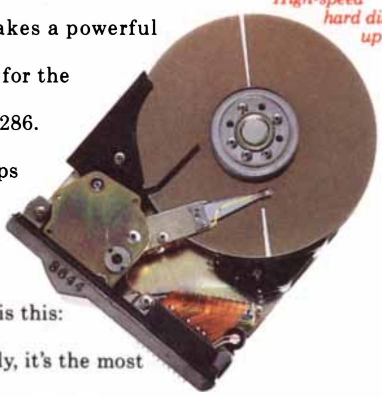


NO-NEUTRINO DOUBLE-BETA DECAY is possible in theory if neutrinos have mass. A neutron inside the nucleus (*a*) would emit a right-handed antineutrino and an electron. Through some rare process (*yellow box*), the antineutrino would transform into a left-handed neutrino (*b*), which is then absorbed by a neutron. The absorption (*c*) would cause the emission of an electron. An observer would see two neutrons turning into two protons and emitting two electrons but no neutrinos.

An awesome display of power.

Undoubtedly, the most important element of a laptop computer is power. After all, if you can't run demanding applications, you might as well run out and demand another computer. Preferably the NEC ProSpeed™ 286 laptop.

Powered by a CMOS 80286 processor, running at a blinding 16MHz, the ProSpeed can accommodate the most sophisticated applications. It's equipped with one megabyte of memory that's expandable to five. And it comes with either 20 or 40 megabyte high-speed hard disk drives. There's even a model with a 100 megabyte drive for data-intensive applications. Now, all this makes a powerful statement for the ProSpeed 286.



But perhaps the most powerful statement is this: quite simply, it's the most powerful 286 laptop you can buy.

ProSpeed is a trademark of NEC Home Electronics, Ltd

C&C Computers and Communications

An awesome power of display.

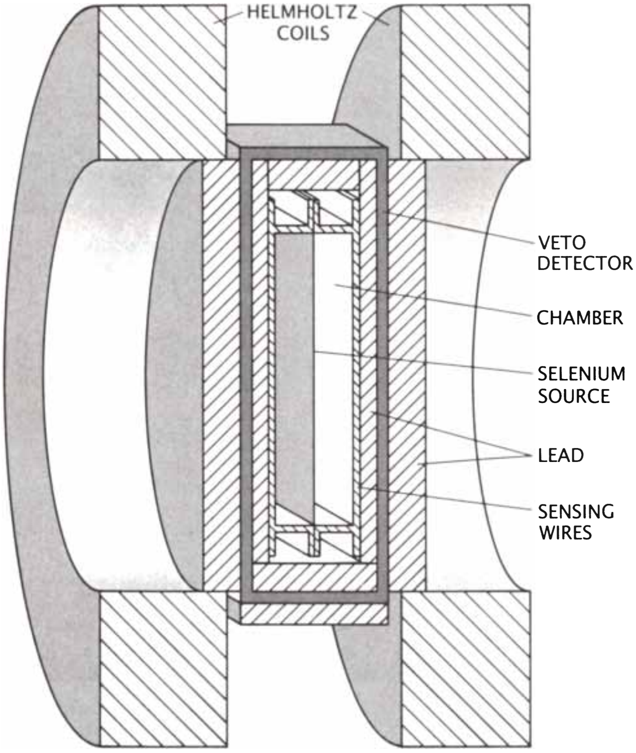
Undoubtedly, the most important element of a laptop computer is its screen. After all, if what you're looking at isn't very clear, then clearly you need another computer. Preferably the NEC ProSpeed 286. Engineered with an advanced screen technology, the ProSpeed gives you crisp, bright, black-and-white images with VGA resolution.

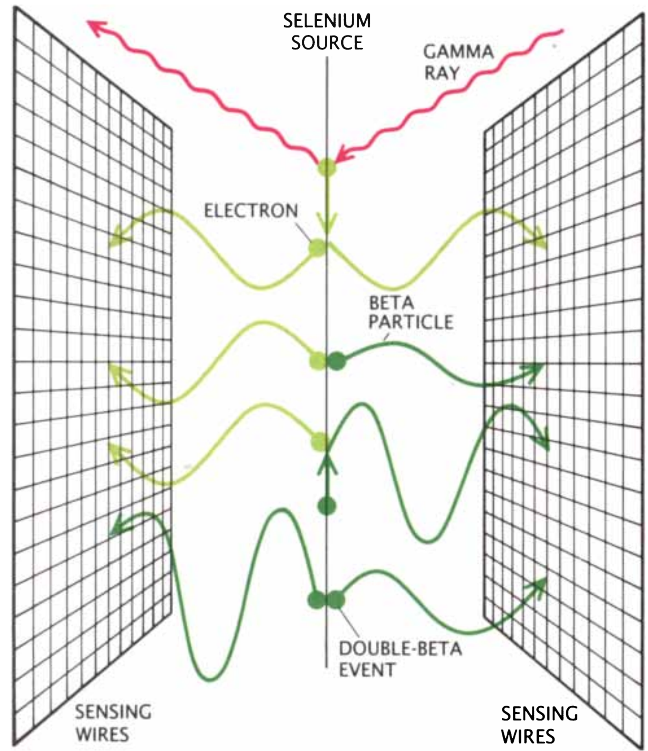
And with more and more people moving towards advanced graphics-oriented applications, the case for the NEC ProSpeed 286 becomes even clearer.

ProSpeed™ 286

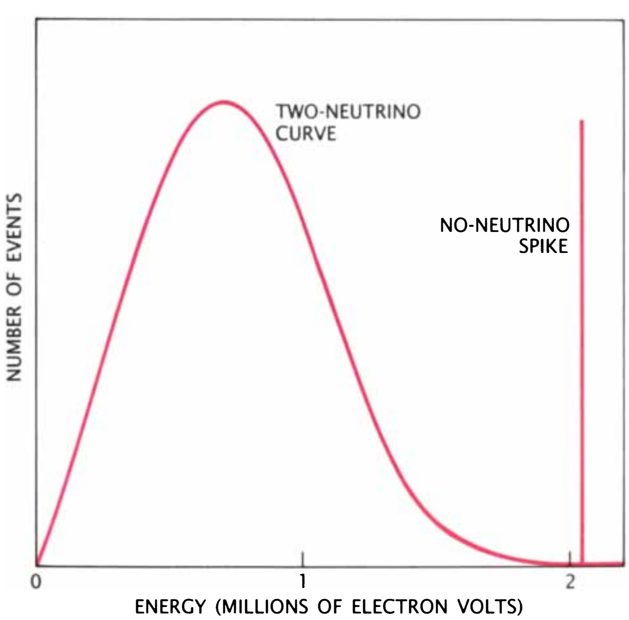
Call 1-800-826-2255 for literature and 1-800-FONE-NEC for technical advice. In Canada, call 1-800-268-3997.

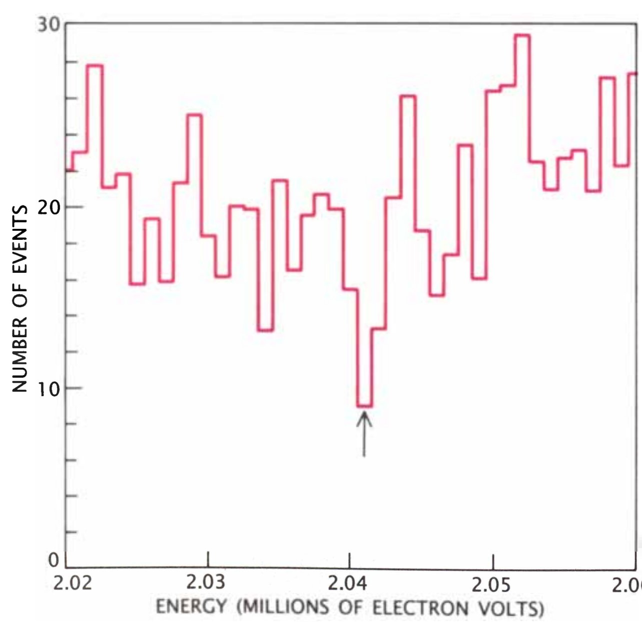
NEC





TIME PROJECTION CHAMBER (*left*) provided the first direct evidence for double-beta decay by tracking the emitted electrons. A sample of selenium 82 is supported in the central plane of the detector. Around the sample is a chamber filled with helium. A lead casing shields the chamber from outside radioactivity, and a "veto" detector warns of incoming cosmic rays. A Helmholtz coil generates a magnetic field, which causes beta rays emitted in the chamber to follow helical paths. As the beta particles move through the helium, they ionize it. An applied electric field causes the resulting free electrons to drift into sensing wires, which register their arrival time and position. The pattern of free electrons is analyzed to recreate the helical paths of the beta rays. The size and pitch of a helix yield the beta-ray energy. The double-beta-decay signature (*bottom event at right*) can be mimicked by the rare background events shown above the signature. Such imposters are usually revealed within a few hours when a daughter nucleus resulting from the event decays at the same spot.





ENERGY SPECTRUM of the electrons associated with germanium 76 decay is expected to include a broad curve for the two-neutrino mode of double-beta decay and a spike for the no-neutrino mode (*left*). The most sensitive measurements to date (*right*) have not revealed the spike, which is predicted to appear at 2.041 million electron volts (*arrow*). The bumpiness in the spectrum is largely the result of statistical fluctuations in the background. If the neutrinoless double-beta contribution is assumed to be less than the size of the statistical fluctuations, the half-life of germanium 76 for the no-neutrino mode must be greater than 2.3×10^{24} years. David O. Caldwell and his co-workers accumulated the data over several years.

did not protect the chamber from cosmic rays, nearly 40 cosmic-ray events penetrated the chamber every second. A cosmic-ray detector was installed to "veto" the recording of such events. Even with cosmic rays largely ignored, the chamber recorded a track about every three seconds.

In this sea of extraneous events, one out of every 90,000 tracks showed the signature of double-beta decay (one track every three days). To help pick out the desired signals, a computer program was written to analyze the data. The workers verified the computer's choices and calculated the energies of the individual electrons.

In order to determine the half-life of double-beta decay and to discover which of the modes had been detected, the group collected data for about one year. At that point enough double-beta-decay candidates had accumulated to give some idea of the energy spectrum—a graph plotting the number of double-beta events against the total energy of the two electrons. For the two-neutrino mode, the energy spectrum was expected to be a broad curve, because for every decay the energy would be distributed differently among the electrons and the neutrinos. For the no-neutrino mode, the electrons would carry off the entire energy of the decay; therefore, the energy spectrum was expected to show a very distinctive spike.

The energy spectrum revealed only the two-neutrino mode; it showed no spike of no-neutrino events. The event rate showed that the half-life of selenium 82 was 1.1×10^{20} years, which means that a billion atoms could survive for a billion years with only a 1 percent chance of a single nucleus succumbing to double-beta decay.

The time projection chamber was recently moved to an underground tunnel at the Hoover Dam to escape the background induced by cosmic rays that evaded the veto detector. The device is still recording double-beta events in search of neutrinoless decay. Although the mode has not been detected, the measurement of the two-neutrino half-life has helped place more restrictive bounds on the mass of the neutrino by supplying one of the factors needed to calculate the mass from the neutrinoless mode.

The no-neutrino decay rate is related to three factors. The first is the mass of the neutrino. The second is the difference in energy between the parent element and its daughter product, which is known as the transition energy. The third is a theoretical quantity called the matrix element, which accounts for interactions between the particles in the nucleus and for the intermediate virtual states. Thus, a measurement of the no-neutrino decay rate could be used to find the neutrino mass, provided that the matrix element could be calculated from fundamental physical principles. But calculations of matrix elements are notoriously difficult, and their values consequently are somewhat uncertain.

In the case of two-neutrino decay, however, the decay rate depends only on the matrix element and the known transition energy. A measurement of the two-neutrino decay rate (or equivalently, the half-life) therefore serves as a direct check on models for the matrix element, giving firmer grounding to one of the factors needed to calculate neutrino mass from estimates of the no-neutrino half-life.

In the hope of ultimately determining the mass of the neutrino, approximately 20 different laboratories worldwide are searching for the no-neutrino mode. The most sensitive detector for measuring neutrinoless decay developed so far is the calorimeter, which measures only one thing: the total energy of the two electrons. The technique—first conceived by physicists Maurice Goldhaber and Edward der Mateosian of the Brookhaven National Laboratory—is effective when the double-beta–emitting isotope and the detector are one and the same. Ettore Fiorini, now at the Italian National Institute for Nuclear Physics, adapted the technique to a germanium-crystal calorimeter, which is a detector containing 8 percent germanium 76, a double-beta parent isotope.

By itself the calorimeter is the most easily fooled of all double-beta detectors, in that it cannot distinguish between double-beta decay and any other process depositing the appropriate energy. Nevertheless, germanium detectors have proved to be very powerful tools in the continuing search for the neutrinoless mode. Single crystals of germanium can be grown very pure and free of most radioactive contamination. A double-beta decay within the bulk of the crystal would be detected with almost 100 percent efficiency. The energy deposited can be measured with such high precision that the search for neutrinoless decay can be confined to a narrow energy "window" centered on the sharp spectral line expected for the neutrinoless mode.

Several crystals can be assembled in a single array, so that the total number of germanium 76 atoms being monitored is substantial. By means of a gamma-ray detector surrounding the germanium, gamma-ray activity in the crystals can often be identified and rejected. David O. Caldwell and his collaborators at the University of California at Santa Barbara and the Lawrence Berkeley Laboratory have constructed the world's most sensitive device, an eight-crystal germanium array. The group has not yet seen an energy spike that would indicate the no-neutrino mode. They conclude that the half-life for neutrinoless double-beta decay in germanium 76 must be at least 2.3×10^{24} years. (Only three or four counts would occur per year in each kilogram of germanium.) This bound for the half-life corresponds to an upper limit for the neutrino's mass of between .6 and 3 electron volts. (By comparison the next lightest known particle, the electron, has a mass of 511,000 electron volts.)

New experiments using crystals enriched to 10 times the natural abundance of germanium 76 are likely to be three to 10 times more sensitive to neutrino mass than present detectors. Large calorimeters and other detectors using molybdenum 100, xenon 136 and other isotopes of high transition energy may eventually surpass even that sensitivity.

The neutrinoless spike at the end of the gracefully curving two-neutrino spectrum has been likened to the pot of gold at the end of the rainbow. Now that the rainbow is in view, nature beckons us even more compellingly toward the horizon, where the no-neutrino mode of double-beta decay may await discovery.

FURTHER READING

DOUBLE BETA DECAY. W. C. Haxton and G. J. Stephenson, Jr., in *Progress in Particle and Nuclear Physics*, Vol. 12, edited by Denys Wilkinson. Pergamon Press, 1984.

DOUBLE BETA DECAY AND MAJORANA NEUTRINO. Masaru Doi, Tsuneyuki Kotani and Eiichi Takasugi in *Progress of Theoretical Physics*, Suppl. No. 83. Kyoto University, 1985.

DIRECT EVIDENCE FOR TWO-NEUTRINO DOUBLE-BETA DECAY IN ^{82}SE. S. R. Elliott, A. A. Hahn and M. K. Moe in *Physical Review Letters*, Vol. 59, No. 18, pages 2020–2023; November 2, 1987.

A REVIEW OF RECENT DEVELOPMENTS IN DOUBLE-BETA DECAY. F. T. Avignone III and R. L. Brodzinski in *Progress in Particle and Nuclear Physics*, Vol. 21, edited by Amand Faessler. Pergamon Press, 1988.

How T Cells See Antigen

On their own, these key actors in the immune response are blind. Other cells must break down foreign material and enfold it in the body's own proteins before displaying it to the T cells

by Howard M. Grey, Alessandro Sette and Søren Buus

The human body is constantly fighting an imperceptible war against invading microbes and malignant cells. The battle is led by the immune system, which can eliminate or neutralize virtually any invader while sparing the body's own tissues at the same time. The main defenders are the white blood cells called lymphocytes, and the counterattack has at least two prongs. The more familiar one consists of the *B* cells, which react to antigen—distinctively foreign material—by secreting antibodies that bind to the invader. Bolstering the activity of the *B* cells, and supplementing it with a second defensive response, are the *T* cells. These lymphocytes help *B* cells to proliferate and secrete antibodies, and they also kill virus-infected and malignant cells directly.

A precise event triggers the immune response: a receptor molecule on the surface of a *B* cell or a *T* cell encounters the antigen to which the cell is programmed to respond and binds to some small part of it, thereby recognizing it as foreign. Aided by other elements of the immune system, the cell then multiplies and fulfills its role as an antibody-secreting *B* cell, a cytotoxic (cell-killing) *T* cell or a helper *T* cell, which secretes substances that mobilize the other cells. *B* cells perform this feat of recognition on their own, interacting with antigens on bacteria or parasites without any intermediary. Yet isolated *T* cells are blind. What more do they need in order to see a foreign substance?

It has become clear during the past several decades that *T* cells have exacting requirements for recognizing antigen. Another kind of cell must act as a so-called accessory cell, chemically processing the antigen and presenting it to the *T* cell in association with certain of the accessory cell's own surface proteins, known as MHC molecules. Immunologists and molecular biologists are still vigorously probing the intricacies of antigen processing, the nature of the MHC molecules and the role they play in presenting antigen to *T* cells. We have already learned much about this key prelude to the immune response, however. What we know promises to lead to new ways of controlling the immune response. It may aid, for example, in the development of synthetic vaccines and of specific therapies for autoimmune diseases such as multiple sclerosis.

HOWARD M. GREY, ALESSANDRO SETTE and SØREN BUUS have been longtime collaborators in the study of antigen processing and presentation. Grey is co-founder and chief technical officer of the Cytel Corporation in La Jolla, Calif., a biotechnology company that is designing immune-modulating drugs. He got an M.D. from New York University in 1957 and has done research at the Scripps Clinic and Research Foundation in La Jolla and at the National Jewish Center for Immunology and Respiratory Medicine in Denver, where until last year he was head of the division of basic immunology. Sette is senior staff scientist at Cytel and assistant professor of immunology at Scripps. He got his Ph.D. from the University of Rome in 1984 and joined Grey at the National Jewish Center in 1986. Buus, who earned an M.D. at the University of Arhus in 1981, is assistant professor at the Institute of Experimental Immunology at the University of Copenhagen.

One of the first indications that *B* and *T* cells see antigen in fundamentally different ways came from work done 30 years ago by P.G.H. Gell and Baruj Benacerraf, who were then at New York University. They found that antibodies (and the cells making them) that were specific for a foreign protein in its normal, intricately folded form often ignored it after it had been denatured—disordered or unfolded. Yet the "cell-mediated" immune response, which is the work of *T* cells, was virtually identical for proteins in their normal and denatured forms. *B* and *T* cells were not known at the time, but these experiments suggested in retrospect that *B* cells and the antibodies they secrete must recognize antigen mainly by its shape, whereas *T* cells respond mostly to its makeup—to the sequence of amino acids in the protein chain, which would be identical regardless of how the molecule was folded.

Subsequently, the evidence mounted that *T* cells respond to antigen only when an accessory cell "presents" it. Macrophages, the immune system's scavenger cells, were the first accessory cells to be identified; later, dendritic cells (specialized cells found in the lymph nodes and spleen), *B* cells themselves and, for some kinds of *T* cell reactions, any nucleated cell in the body were added to the list. It turned out that the activity of accessory cells, or antigen-presenting cells (APC's), explains why *T* cells have no interest in antigen shape: the APC's break down the antigen before presenting it, obscuring its shape and leaving only its distinctive amino acid sequence.

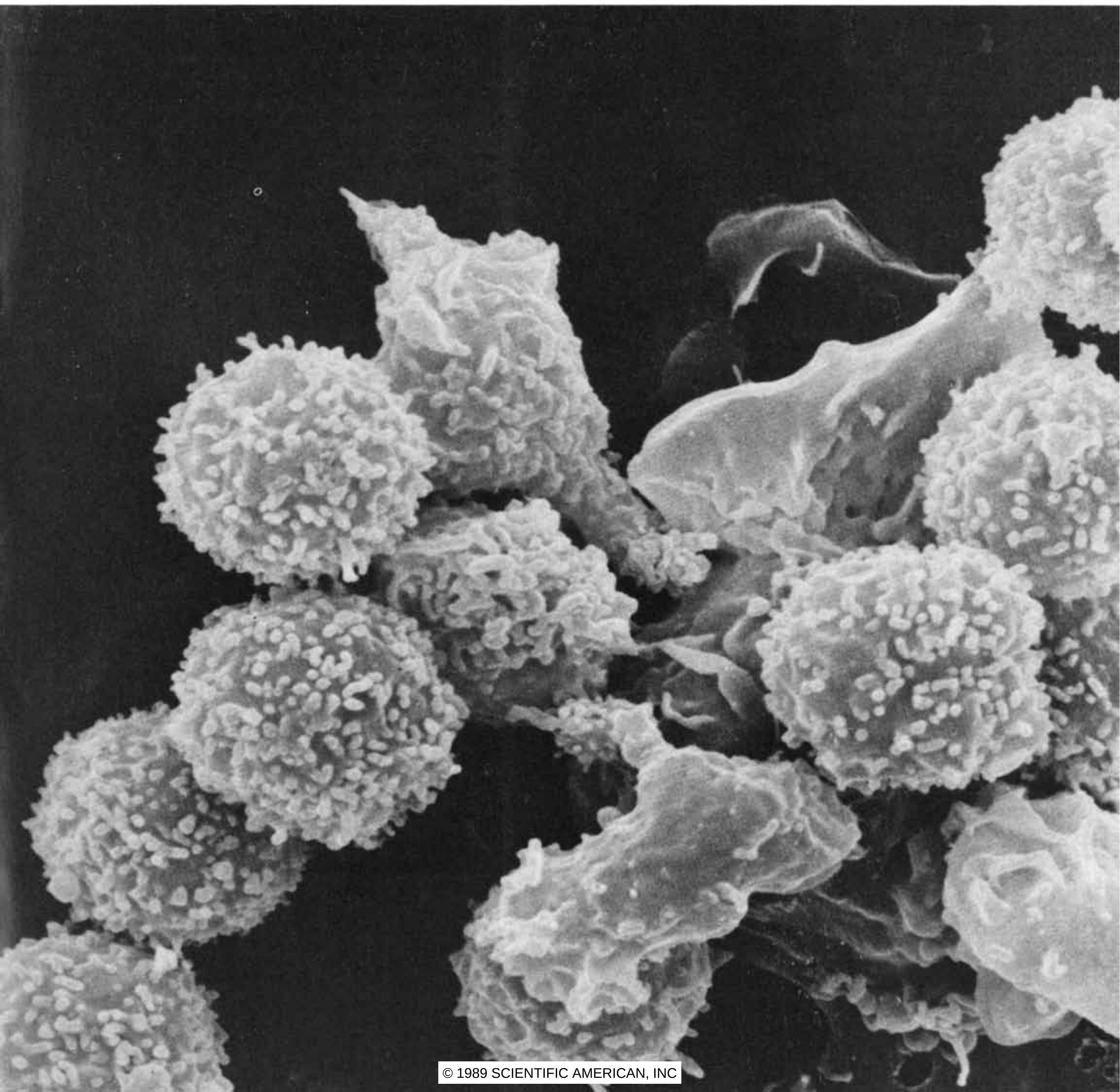
Several studies showed that APC's do more than simply capture antigen and display it on their surface. A technique introduced in 1981 by Emil R. Unanue, then at Harvard University, yielded the most compelling results. He and his colleagues exposed APC's

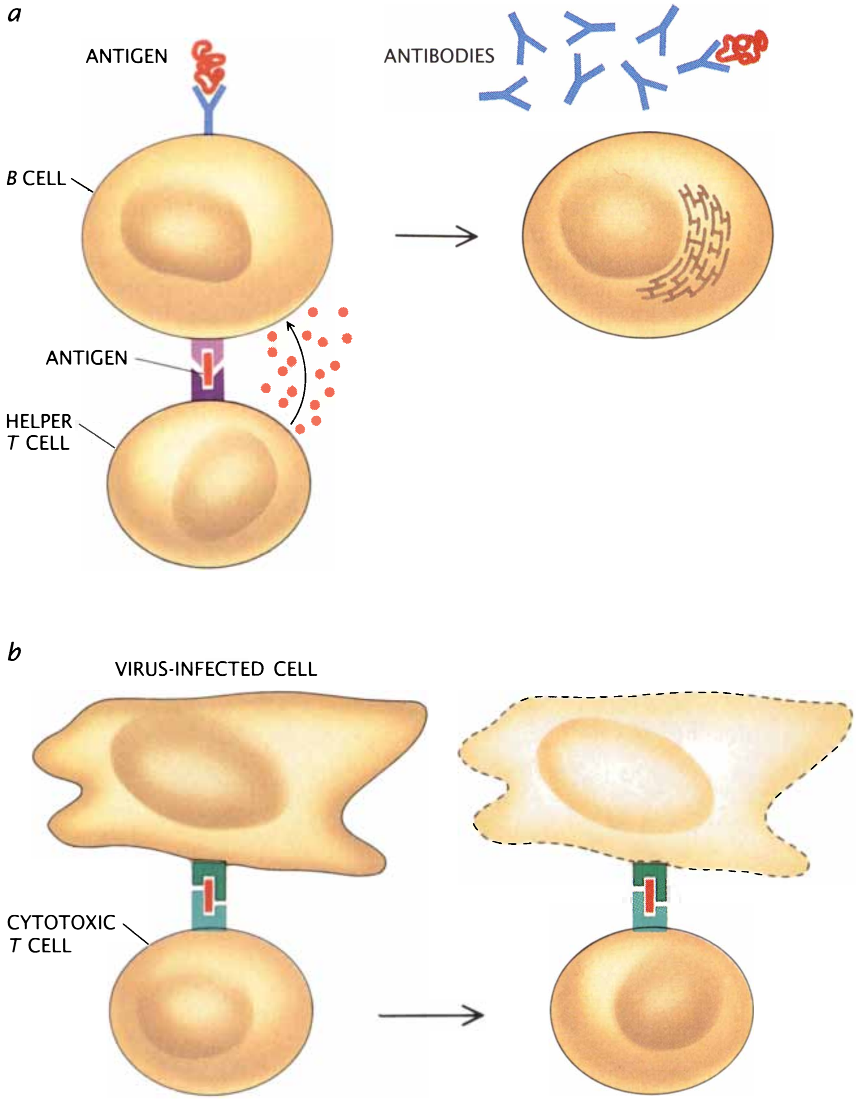
_T_ CELLS RECOGNIZE ANTIGEN on the surface of a macrophage, a scavenger cell. The flat macrophage has ingested a bacterial protein, broken it down and displayed the pieces together with certain of the cell's surface proteins. The small, spherical _T_ cells are programmed to recognize the bacterial antigens but can only do so with the macrophage's assistance. The scanning electron micrograph is by Morten H. Nielsen and Ole Werdelin of the University of Copenhagen.

to antigen and then, after varying intervals, "fixed" the cells with formaldehyde, which interrupted their metabolism. The workers then tested the cells' ability to present antigen to *T* cells and trigger their proliferation. APC's fixed before or immediately after they were exposed to antigen could not present it to *T* cells. In contrast, APC's that were incubated with antigen for an hour or more and then fixed were perfectly capable of doing so. These results and others suggested that, after exposure to antigen, accessory cells required time and energy before they could present antigen to *T* cells, probably because they had to modify it somehow first.

Other experiments shed light on this process by showing that certain weak bases strongly inhibit the ability of APC's to present antigen. The compounds are probably active in the endosomes, acidic compartments within the cell where ingested material is broken down by proteolytic, or protein-cleaving, enzymes. Presumably by neutralizing the endosomes, the bases inhibit a cell's ability to degrade proteins. Later work showed that specific inhibitors of proteolytic enzymes also interfere with antigen presentation.

The possibility that cleavage of antigen into short fragments, or peptides, prepares it for presentation to *T* cells gained crucial support from an experiment done by Richard P. Shimonkevitz, Philippa C. Marrack and John W. Kappler, all of the National Jewish Center for Immunology and Respiratory Medicine in Denver, and one of us (Grey). The group showed that single peptides derived from a protein antigen could substitute for the intact protein in triggering a *T* cell response. The peptides clearly needed no further processing to do so, since they could be presented by APC's that had





T CELLS HAVE A DUAL ROLE. A helper *T* cell recognizes antigen (*color*) on the surface of another cell (in this case a *B* cell) that has encountered foreign material and broken it down (*a*). The *T* cell responds by secreting substances that help other immune-system cells to fulfill their roles. (Here it stimulates the *B* cell to mature and secrete antibodies.) A cytotoxic, or killer, *T* cell recognizes antigen on the surface of a virus-infected or malignant cell and responds by killing the target cell (*b*).

previously been fixed. Recently Stephane O. Demotz in our laboratory at Cytel actually isolated a processed antigen and determined that it is indeed a short peptide. In one account of antigen processing, then, an APC engulfs antigen and delivers it to acidic compartments within the cell, where it is broken down into small peptides, as short as 10 to 20 amino acids, before being returned to the cell membrane for recognition.

That is only a partial account of antigen processing, however. The steps it describes occur in the specific classes of antigen-presenting cells—*B* cells, macrophages and dendritic cells—that are specialized for processing foreign material taken in from the surrounding medium. The processing and presentation of such "exogenous" antigens generally leads to the activation of a specific population of *T* cells: the helper cells that aid *B* cells in making antibody.

Not all the antigens recognized by *T* cells originate outside the presenting cells, however. A cell that has been infected by a virus or has become malignant may synthesize distinctive, virus- or tumor-specific proteins. Virtually all cells in the body can present such internally synthesized proteins, and they do so to *T* cells belonging to the second major population: the cytotoxic *T* cells. These lymphocytes respond to "endogenous" antigens by killing the cells that produce them.

Until recently many workers assumed that such endogenous antigens did not need to be processed, since the intact proteins are often expressed on the surface of the abnormal cells. It seemed plausible that cytotoxic *T* cells, unlike helper cells, might be able to respond directly to intact antigen. Yet Alain Townsend of the John Radcliffe Hospital in Oxford found in 1985 that cytotoxic cells capable of killing cells infected with a virus could also kill uninfected cells into which a mere fragment of a viral gene had been introduced. These genetically engineered cells produced only a small fraction of the corresponding viral protein found on the infected cells. The cytotoxic cells nonetheless responded identically to both molecules. Later Townsend induced a cytotoxic response with uninfected APC's that had merely been incubated with a short antigenic peptide, confirming that cytotoxic cells, like helper cells, recognize a fragment of antigen and not the complete protein.

Other experiments indicated, however, that infected cells process endogenous antigens by a mechanism quite different from the one that prepares antigen for recognition by helper cells. The weak bases that blocked processing for helper *T* cell recognition and pointed to a central role for endosomes in that processing pathway had no effect on antigen presentation to cytotoxic *T* cells. Moreover, when an antigenic protein was added to a culture containing APC's and antigen-specific cytotoxic *T* cells, nothing happened. The killer cells were able to recognize and respond to the protein, however, when it was microinjected into the cytoplasm—the fluid interior medium—of the presenting cells.

The data are not yet completely definitive, but they are most compatible with a picture in which endogenous antigen is processed in the cytoplasm rather than within endosomes. Once the protein has been degraded in the cytoplasm, the fragments are somehow moved into the interior of a vesicle, a sac that shuttles between the cell interior and its surface. The peptides are then transported to the cell surface for recognition by killer *T* cells.

This second processing pathway, specialized for antigens made by the APC itself, could be the immune system's way of ensuring that a foreign organism cannot elude it by adopting a Trojan-horse strategy. Even if the pathogen is hidden within a cell, the body will process the novel proteins and make them visible to the *T* cells.

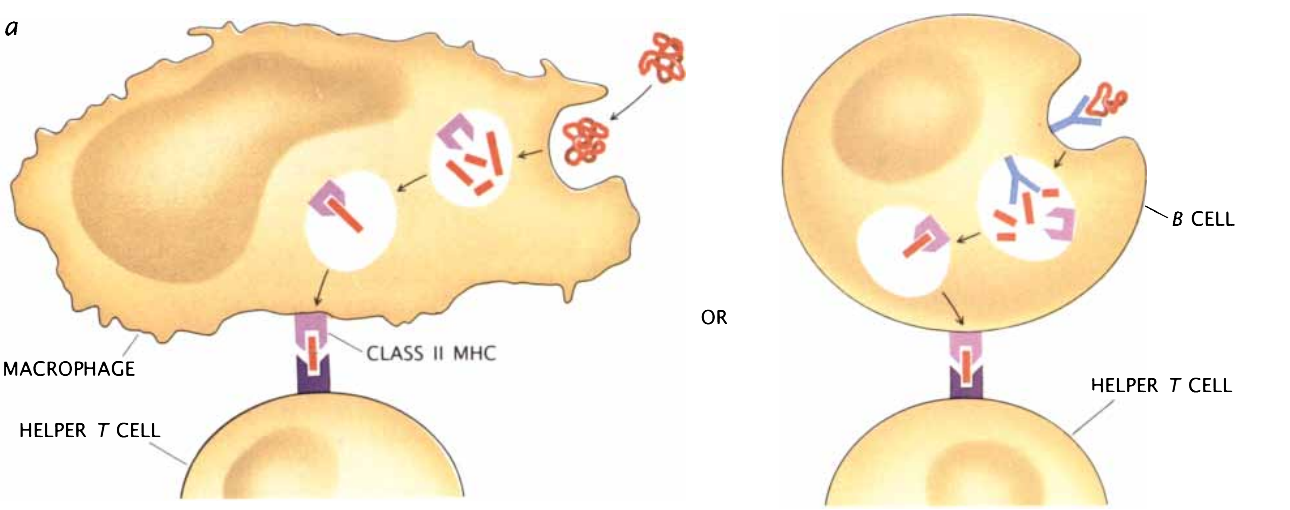
In addition, the existence of two separate pathways of antigen processing, one for exogenous antigens and one for endogenous antigens, makes biological sense: each pathway leads to the appropriate *T* cell response. A bacterial protein taken up by a *B* cell from its surroundings and processed by the exogenous pathway elicits *T* cell help, which enables the *B* cell to produce antibodies for combating the infection. A foreign or abnormal protein made by a renegade cell, in contrast, leads to the killing of the errant cell by cytotoxic *T* cells.

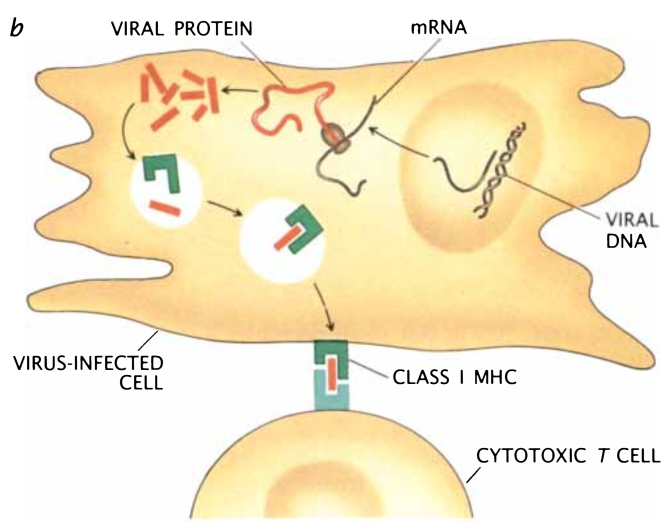
Once it has been processed, antigen is displayed on the surface of the accessory cell together with proteins of the cell's own making. They are known as MHC proteins, after the major histocompatibility gene complex, a cluster of more than a dozen genes. The cluster is a hot spot of genetic variability, so that the MHC proteins encoded by a given set of genes almost always differ from one individual to the next. The molecules do fall into two broad classes, however, according to their structure and their role in *T* cell stimulation. Class II MHC proteins, found mainly on the surface of *B* cells, macrophages and dendritic cells, figure in the presentation of antigen to helper *T* cells. Class I proteins, found on almost all nucleated cells in the body, play the same role for cytotoxic *T* cells.

The current picture of MHC molecules and their part in stimulating the *T* cell response is the product of more than three decades of investigation, beginning in the mid-1950's with studies of tissue grafts. Investigators found that when tissue from one animal was transferred to another one with different MHC proteins, the immune system of the recipient rejected the graft in an extraordinarily intense reaction, one that was later traced to *T* cells. It appeared that the immune system, and its *T* cell arm in particular, is "tuned" to recognizing MHC molecules. Clearly, though, their normal immunologic function had to be something other than graft rejection. After all, grafts are rare in nature.

An early hint of a normal function for MHC proteins came from experiments done in the 1960's by Hugh O. McDevitt, who was then at the National Institute for Medical Research in England, and Benacerraf. They showed that the genes of the MHC affected an animal's ability to mount an immune response to certain simple antigens. An animal carrying one variant of a particular MHC gene might respond vigorously to a given antigen; another animal carrying a different variant might not respond at all. In these responder and nonresponder strains, the MHC seemed to function as "immune response" genes.

How might these genes affect the immune response? The most obvious explanation was that they encoded the *T* cells' own receptor molecules. In 1973, however, Alan S. Rosenthal and Ethan M. Shevach of the National Institute of Allergy and Infectious Diseases made an observation that linked the MHC to the function of the accessory cells. They mated a guinea-pig strain





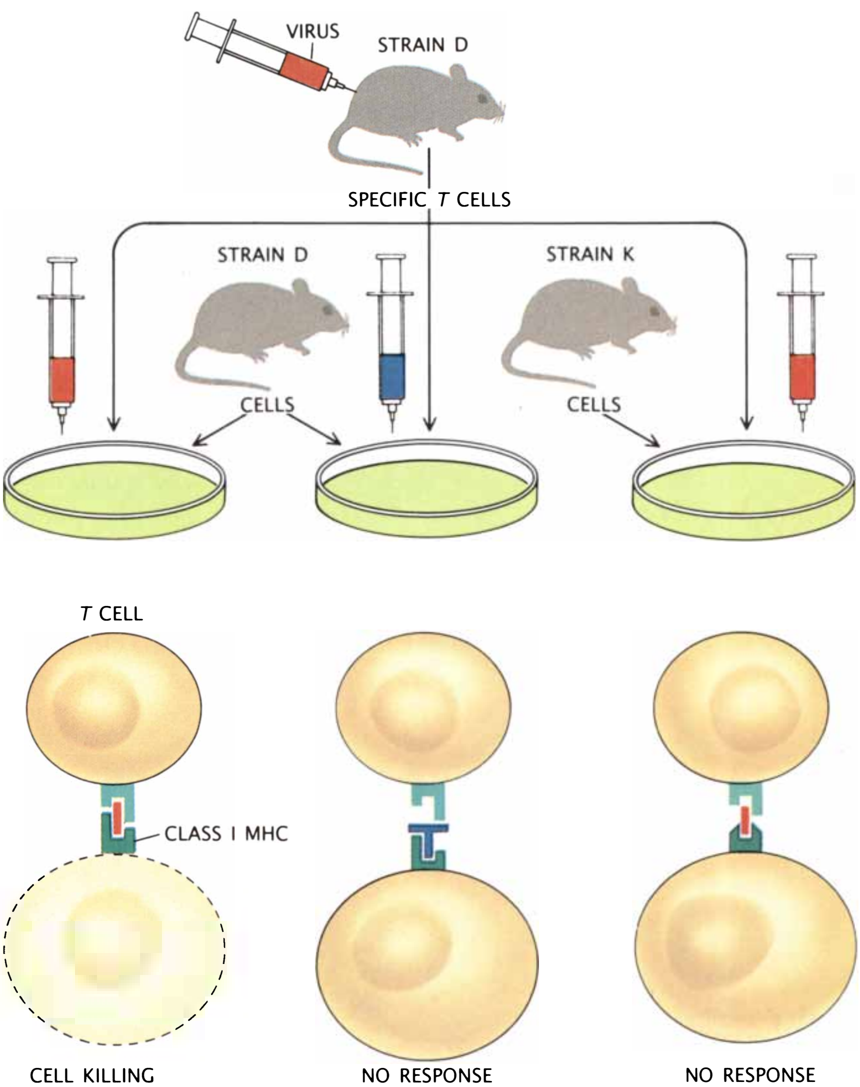
ANTIGEN IS PROCESSED by one of two pathways, depending on its origin. Foreign material from outside the cell (material from bacteria or parasites, for example) is engulfed by a specialized cell (*a*), often a macrophage (*left*) or a *B* cell (*right*). The cell breaks down the material in an intracellular compartment and links the processed antigen to class II MHC molecules. The complexes are transported to the cell surface and displayed there to helper *T* cells. Abnormal proteins made within a virus-infected or malignant cell (*b*) are broken down in the cell's internal medium. Only then do the fragments move into a vesicle, or sac, where they join class I MHC molecules and are carried to the cell surface for display to cytotoxic *T* cells.

that responded well to one antigen and poorly to a second antigen with a strain that showed the reciprocal pattern, responding poorly to the first antigen and well to the second. The offspring—having inherited a gene for responsiveness from each parent—could mount a strong response to both antigens. But when the workers extracted *T* cells from the hybrid animals and mixed them with APC's and antigen in culture, the helper *T* cell response depended on the origin of the accessory cells.

In the presence of APC's that had also come from the hybrid animals, the *T* cells responded to both antigens, as expected. When the APC's had been isolated from the parental strains, however, the *T* cells reacted only to the antigen to which the parental strain had also responded. The cells seemed blind to the other antigen. It appeared that the MHC genes exerted their effect not through the *T* cells themselves (under the right conditions the *T* cells of the hybrid animals were perfectly capable of responding to both antigens) but through the APC's. Somehow accessory cells carrying the nonresponder genes were unable to present one of the antigens—a phenomenon for which the mechanism has only recently been elucidated.

Meanwhile studies of cytotoxic *T* cells led to the conclusion that *T* cells recognize not only foreign antigen but also the MHC-encoded proteins on the accessory cells. In 1974, for example, Rolf M. Zinkernagel and Peter C. Doherty of the Australian National University exposed *T* cells that had responded to antigen presented by cells carrying a particular variant of class I MHC protein to the same antigen presented by cells bearing a different MHC variant. First, the workers infected a mouse with a virus, stimulating cytotoxic *T* cells targeted to the virus-infected cells. Then they extracted the specific *T* cells and exposed them in vitro to virus-infected cells from other mice.

Zinkernagel and Doherty found that if the class I MHC proteins on the surface of these new infected cells differed from those of the original mouse, the cells escaped *T* cell killing. The workers interpreted the results as showing that an animal's *T* cells had to recognize two entities in order to respond: both an antigen and a specific MHC protein—one that is characteristic of the animal's own cells. Confirmed by many other experiments, this requirement for co-recognition of antigen and a "self" MHC molecule became known as MHC restriction.



MHC RESTRICTION of the *T* cell response, discovered by Rolf M. Zinkernagel and Peter C. Doherty of the Australian National University in 1974, consists in *T* cells' need to recognize both a specific antigen and a specific MHC protein. The workers infected mice of a specific MHC strain with a virus and isolated virus-specific cytotoxic *T* cells. In culture, these lymphocytes could kill cells from uninfected mice of the same strain when the cells were infected with the same virus (*left dish*) but not when the virus was different (*center dish*). They also failed to kill cells from another strain carrying the same virus but bearing a different MHC protein (*right dish*).

This MHC restriction of the *T* cell response presented a new puzzle. *B* cells, after all, are activated by the fit of a single key (the antigen) into a single lock (the receptor on the *B* cell surface). What might be the molecular design of the *T* cell's double-key system? One theory held that *T* cells bear two independent receptor molecules, one specific for antigen and the other for a self-MHC protein. A second theory postulated that *T* cells carry a single receptor molecule capable of identifying both antigen and MHC. Proponents of each theory raised indirect evidence in its favor; the controversy was settled in favor of the one-receptor model when a single *T* cell receptor was shown to be specific for both antigen and self-MHC [see "The *T* Cell and Its Receptor," by Philippa Marrack and John Kappler; SCIENTIFIC AMERICAN, February, 1986].

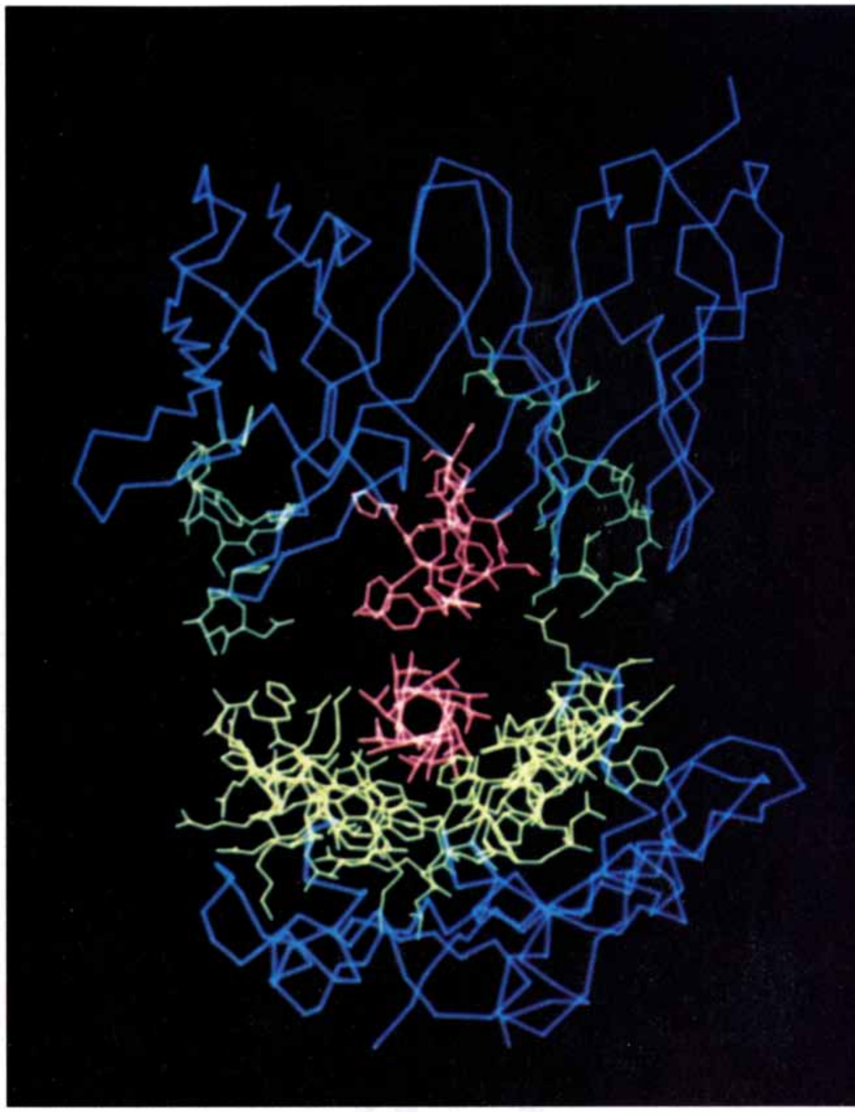
The existence of a single receptor suggested that the processed antigen and the MHC molecule might form a complex—a single entity that would fit a single recognition site in the *T* cell receptor. In effect, the MHC protein would act as the primary receptor for processed antigen; the resulting complex would then interact with a second receptor, on the *T* cell. Because both the antigen and the MHC would contribute to shaping the molecular characteristics of the complex, the proposed mechanism would elegantly explain *T* cells' specificity for both MHC and antigen. It might also explain the puzzle early studies had posed: How do certain MHC genes render an individual blind to specific antigens? In this new picture, those genes might encode proteins unable to bind and present particular peptides.

Ronald H. Schwartz of the National Institute of Allergy and Infectious Diseases provided compelling but indirect evidence in favor of complex formation. He studied the ability of mice belonging to different MHC strains to react to variants of a particular protein. He found that whereas a specific variant might elicit a *T* cell response in one strain but not in a second one, a difference of a few amino acids in the protein's sequence might make it visible to the immune system of the second strain. Schwartz argued that such results were best explained by supposing that the protein—or a peptide cleaved from it—had to bind to MHC molecules before it could trigger a response. The slight difference in amino acid sequence was what was needed for the peptide to bind to the MHC molecules of the second strain.

In 1985 Unanue and his colleagues at Washington University were the first to demonstrate complex formation directly, by means of a technique called equilibrium dialysis. A chamber containing an antigenic peptide was separated by a semipermeable membrane from another chamber containing the class II MHC molecule that restricted the immune response to the antigen. The antigen—by far the smaller molecule—could pass through the membrane, but the MHC protein was confined on one side. All else being equal, the antigen should have diffused through the membrane until its concentration in both chambers was the same. Instead its concentration grew larger on the side that also contained the MHC protein. Evidently, the molecules were binding to each other.

Our group demonstrated the same kind of interaction for a variety of peptides and class II MHC molecules. We also showed that the binding is critical for an immune response: *T* cells recognize complexes of MHC and antigen. Adopting a technique developed in Harden M. McConnell's laboratory at Stanford University, we embedded antigen-MHC complexes into an artificial lipid membrane—a simulated cell membrane. For comparison, we also prepared membrane containing uncomplexed MHC, bathed in free antigen. The preformed complexes stimulated antigen-specific *T* cells some 20,000 times more efficiently than did the uncomplexed MHC and antigen.

For such complexes to have a role in the normal immune response, they must be quite stable: in any individual only a few *T* cells bear receptors specific for a given antigen, so that after an individual's exposure to the antigen it may take some time before a specific *T* cell encounters an APC bearing the antigen-MHC complex. The success of our experiment suggested that the complexes are indeed stable, since it took us more than a day to isolate antigen-MHC complexes and embed them in the lipid membrane. Direct measurements of the complexes' dissociation rates confirmed their stability: at body temperature their half-life was about 10 hours.



CO-RECOGNITION of antigen and an MHC protein is modeled. A *T* cell's receptor approaches from above and encounters a foreign peptide (*pink pinwheel*) in the antigen-binding cleft (*yellow trough*) of the MHC protein. *T* cell receptors include fairly constant regions (*yellow*), which interact with the body's small range of MHC proteins, and a highly variable region (*pink*), for recognizing diverse antigens. The computer model is by Mark M. Davis and Pamela J. Bjorkman of Stanford University.

Compelling as the evidence of complex formation was, not everyone accepted the further proposal that a failure of some MHC proteins to bind certain antigens underlies the genetic unresponsiveness investigators such as Rosenthal and Shevach had studied. Experiments with different antigens did not always support Rosenthal and Shevach's conclusion that such immunologic blind spots reflect a deficit in the antigen-presenting cells. Also, some workers pointed out that it was hard to see how a single MHC protein could act as a specific receptor for myriad structurally distinct antigenic peptides. After all, each individual has at most about a dozen MHC proteins. How could the MHC proteins be selective when each one must bind a sizable part of a vast universe of potential antigens? In this view, antigen-MHC complexes, if they existed, had to form nonspecifically. Variations in immune responsiveness had to reflect something other than selective binding.

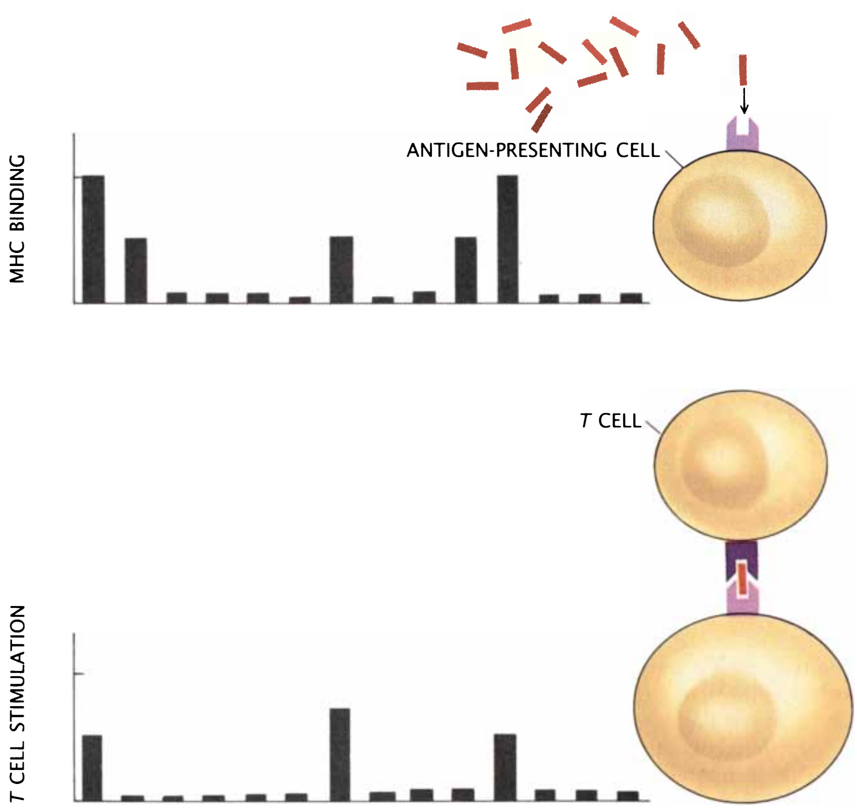
Some investigators proposed instead that the MHC influences the immune response by shaping the repertoire of functional *T* cells. *T* cells mature in the thymus gland, and in the process they interact with the MHC proteins on the surface of accessory cells in the thymus. During this thymic "education," the *T* cells learn to recognize antigen only in association with the body's own MHC molecules. At the same time, it is thought, *T* cells that bind too avidly to self-MHC—and hence pose the threat of an autoimmune reaction—are eliminated or at least inactivated. Conceivably, a particular variant of a self-MHC protein might lead to the elimination of all the *T* cells capable of reacting to a particular antigen. Any individual inheriting the corresponding MHC gene would display the same hole in the *T* cell repertoire.

We tested the relative influences on immune responsiveness of MHC binding and holes in the *T* cell repertoire by comparing peptides' ability to bind to a mouse MHC molecule with their ability to induce an immune response. Of a set of 14 peptides—which in sum represented an entire protein molecule—five could bind to the MHC protein; three of those five, we found, could then trigger a *T* cell response in animals of the same MHC strain. None of the peptides that failed to bind could stimulate a response.

The selectivity of the MHC proteins, then, does shape the immune response. But not every peptide that can bind to a self-MHC elicits a response; some antigens that bind fail to stimulate a response, apparently because *T* cells able to recognize the antigen-MHC complex are absent. Both theories of how the MHC genes influence the immune response appear to be correct. The selectivity of MHC proteins in binding antigens combines with holes in the *T* cell arsenal to set the boundaries of an individual's immune responsiveness.

The earlier objection to the notion of MHC proteins as specific antigen receptors remained unanswered, however: How could an MHC protein selectively bind many—but not all—antigens? We found that a typical MHC molecule can indeed bind between 10 and 20 percent of the peptide fragments from any given protein molecule. We also identified a possible basis for this broad but selective binding: peptides bound by a particular MHC molecule turned out to share certain simple structural features.

One MHC molecule, for example, bound peptides that all shared a motif of repeated hydrophobic residues—amino acids with an affinity for a nonwater medium. Another MHC molecule bound peptides that had in common a trio of positively charged residues. Perhaps such diverse, broad specificities give an individual's array of MHC proteins the ability to bind and present the widest possible variety of antigens, so that a foreign substance is unlikely to slip through the immune system's defenses.



IMPORTANCE OF MHC BINDING to an antigen's ability to stimulate a *T* cell response was assessed by the authors. They synthesized 14 peptides representing fragments of a protein and measured the affinity of each one for a mouse MHC molecule (*top*). Five of the peptides bound to the molecule. Three of the five could stimulate a *T* cell response in a mouse of the same strain (*bottom*). Binding to an MHC protein appears to be necessary, but not sufficient, for a peptide to trigger an immune response.

Vivid confirmation that MHC molecules serve as receptors for processed antigen exported to the cell surface came in 1987, when Don C. Wiley and his colleagues at Harvard University solved the three-dimensional structure of a class I MHC molecule. The most striking feature of the structure, determined from the diffraction pattern of X rays trained on a crystal of the protein, was a cleft on the top of the molecule, where it would face outward from the cell surface. Two helical regions of the protein form the walls of the cleft; so-called beta sheets, in which the protein chain folds back and forth in a plane, form the cleft's floor.

The cleft *looks* like the binding site for an antigenic peptide. What is more, many of the variable amino acids that distinguish a particular MHC protein in different individuals and affect immune responsiveness turn out to be clustered on the inside walls and floor of the cleft. Such amino acids presumably influence the protein's peptide-binding ability; one would therefore expect them to mark the binding site.

A second observation also pointed to the cleft as the peptide-binding site and raised an intriguing new possibility about the function of MHC molecules. The cleft was not empty: in it Wiley and his colleagues identified another molecular entity. The material must have been bound to the MHC molecules when they were crystallized; presumably, it was a piece of processed antigen.

Paul M. Allen of Washington University and our own group have confirmed that the binding site on MHC proteins of accessory cells is routinely occupied. Acid treatment of class II MHC molecules purified from *B* cells released peptides that later could rebind specifically to the MHC molecules. What is more, Townsend and his co-workers recently showed that a cell cannot even assemble class I MHC molecules properly unless a peptide is present during the final stages of the protein's folding process. It seems likely that these omnipresent peptides are fragments of the body's own proteins, produced within the cell or captured from its surroundings, that have been processed and presented by the same mechanisms that display foreign antigens.

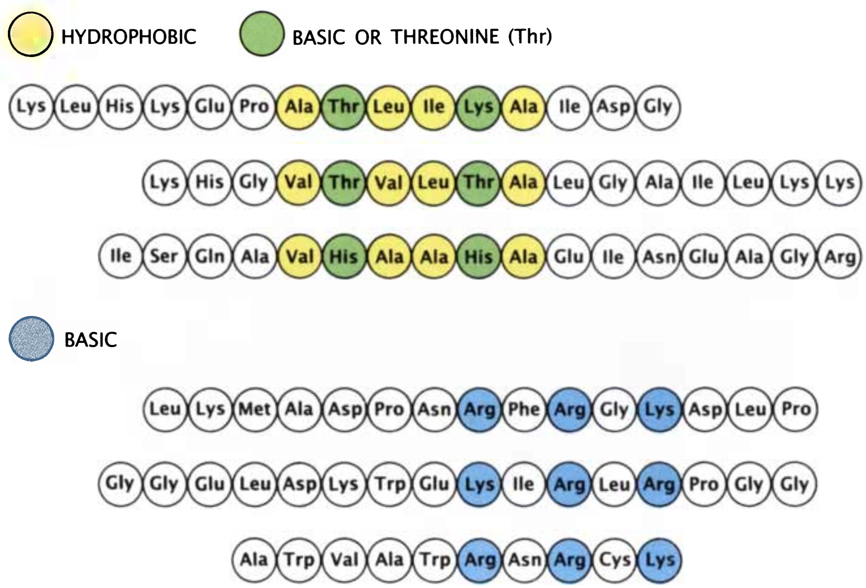
That proposal is consistent with the theory of immune surveillance, which holds that killer *T* cells constantly monitor the other cells of the body for the appearance of tumor or viral antigens and promptly eliminate any cell expressing them. By continuously processing and presenting their own antigens, cells in effect invite inspection by the immune system, so that it can quickly detect any aberration.

This scenario of constant self-scrutiny suggests an answer to the inevitable question about antigen presentation to *T* cells: Why does it need to be so elaborate? Why do *T* cells not recognize antigen directly, as *B* cells do, instead of requiring it to be broken down and displayed in the context of MHC molecules? For cytotoxic *T* cells, one answer is that MHC restriction targets them to the body's own tissue, where—being killer cells—they are designed to act. Because the cells are "interested" in self-MHC as well as antigen, they look for antigen in precisely the setting in which they can respond effectively to it.

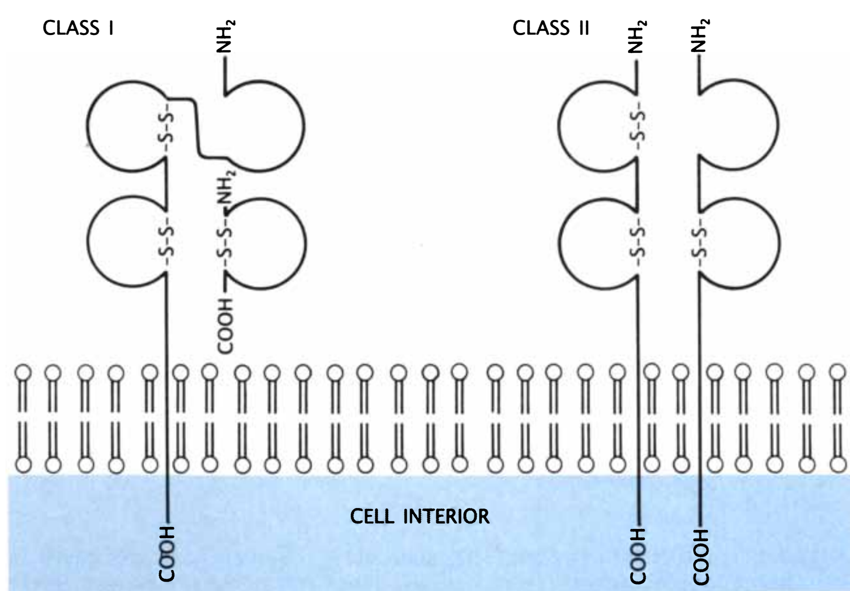
For the equally complex scheme of antigen presentation to helper *T* cells, one might invoke an evolutionary explanation. Cell-mediated immunity appears to be ancient; even organisms as primitive as sponges can recognize and prevent invasion by cells from different species. Thus, *T* cells may have originated as killer cells, but even when they acquired an additional, helper role, they retained a disposition to look for antigen on the cell surface, in association with the body's own proteins. Over the course of evolution, this interest in self-proteins became adapted to the helper cells' function, so that class II MHC proteins now guide these *T* cells to a site where they can be most effective—to *B* cells, the primary target of *T* cell help.

Even though the picture of antigen processing and presentation is far from complete, understanding of the phenomenon has advanced dramatically. The result is likely to be an improved ability to manipulate the immune system for clinical purposes: to stimulate immunity with vaccines and selectively suppress it in autoimmune diseases.

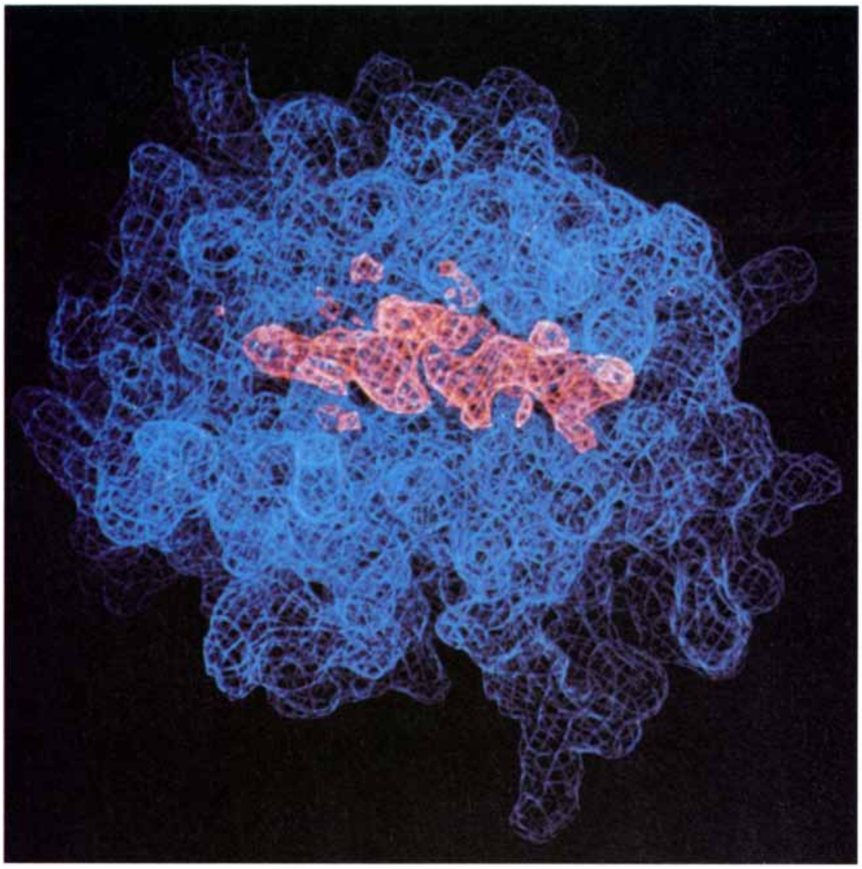
Traditionally, vaccines have consisted of the whole pathogenic organism, live or killed, or a protein extracted from it. For some diseases, such as malaria, that approach is not feasible, and some whole-organism vaccines

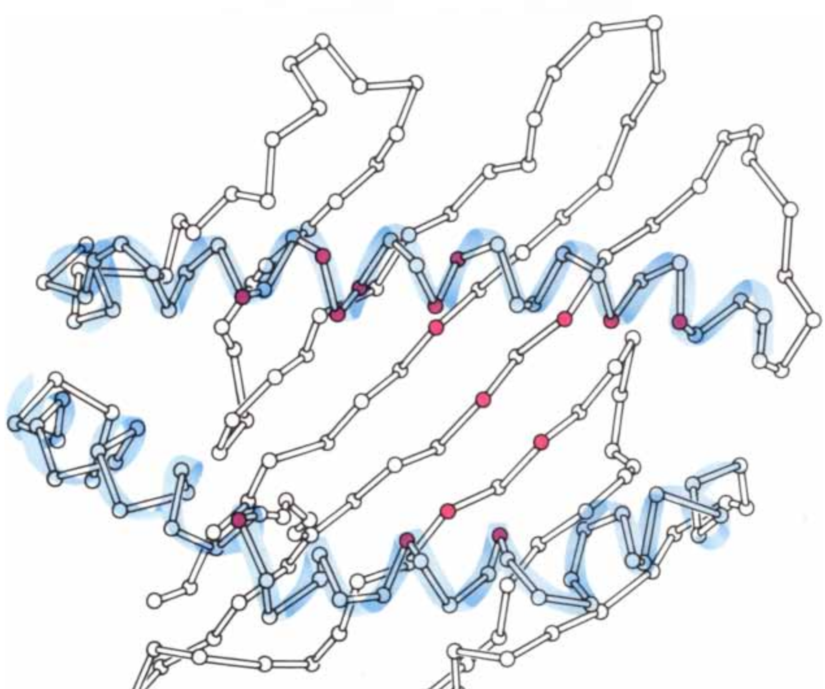


DISTINCTIVE STRUCTURAL MOTIFS characterize peptides able to bind to particular MHC proteins. The upper three peptides all bind well to one MHC molecule; the lower three all bind well to another such molecule. Peptides in each cluster share a common pattern defined by the chemical properties of their constituent amino acids.



MHC MOLECULES are divided by structure into class I, found on almost all cells, and class II, found only on specialized antigen-presenting cells. Each protein consists of two chains. In class I molecules one chain extends through the cell membrane and the other lies outside the cell; in class II molecules both chains extend through the membrane. Bonds between sulfur atoms (*S*) divide the chains into looping domains. On proteins of both classes the binding site for antigen lies within the upper loops.





ANTIGEN-BINDING CLEFT of a class I MHC molecule is shown in two images based on X-ray analysis of the protein: a computer model (*top*) and a diagram (*bottom*), where the cleft lies between the two helixes. The discovery of a separate substance (colored orange in the computer image) lodged in the cleft supported the proposal that it is the binding site for antigen. In addition, many of the variable amino acids (shown in red in the diagram) that affect a particular MHC molecule's antigen-binding capability are clustered in the cleft. Don C. Wiley of Harvard University and his colleagues Pamela J. Bjorkman, Mark A. Saper, Boudjema Samraoui, William S. Bennett and Jack L. Strominger determined the molecular structure and provided the computer image.

have risky side effects. In such cases vaccine developers are now trying to design synthetic peptides (representing only a small part of the actual antigen) that will trigger an equivalent immune response. To do so, the peptides must stimulate helper and cytotoxic *T* cells as well as *B* cells, and so it is critical that these antigens bind to MHC molecules in spite of individual variation. The deepening understanding of MHC-antigen interactions will surely help guide the design of such peptide vaccines.

It may also help in treating insulin-dependent diabetes, rheumatoid arthritis and multiple sclerosis—diseases in which the immune system loses its ability to discriminate self from nonself and responds to the body's own molecules. Some of these diseases almost exclusively affect people carrying specific MHC genes. The corresponding MHC proteins may play a role in the diseases by presenting self-antigens in a way that induces an immune response.

MHC-based technology should make it possible to develop compounds that bind very strongly to the disease-associated MHC proteins. By blocking the binding of self-antigens, such compounds might suppress the autoimmune response. It is already possible to arrest some of these diseases by means of an immunosuppressive agent (such as cyclosporine), which blocks the immune response generally. A blocker targeted to a specific MHC variant, however, would have the advantage of leaving the immune system largely intact and able to defend the body against external threats. Growing understanding of antigen processing and presentation may thus give the fight against autoimmune diseases some of the immune system's own precision and power.

FURTHER READING

ANTIGEN PRESENTING FUNCTION OF THE MACROPHAGE. Emil R. Unanue in *Annual Review of Immunology*, Vol. 2, pages 395-428; 1984.

ANTIGEN PRESENTATION PATHWAYS TO CLASS I AND CLASS II MHC-RESTRICTED T LYMPHOCYTES. Thomas J. Braciale et al. in *Immunological Reviews*, No. 98, pages 95-114; August, 1987.

THE INTERACTION BETWEEN PROTEIN-DERIVED IMMUNOGENIC PEPTIDES AND IA. Søren Buus, Alessandro Sette and Howard M. Grey in *Immunological Reviews*, No. 98, pages 115-141; August, 1987.

STRUCTURE OF THE HUMAN CLASS I HISTOCOMPATIBILITY ANTIGEN, HLA-A2. P. J. Bjorkman et al. in *Nature*, Vol. 329, No. 6139, pages 506-512; October 8, 1987.


Liqueur
Triple Orange
Triple Orange
Grand Marnier
Liqueur
PARIS
FRANCE

and controls are within easy reach. And to make sure you're always in complete control, Talon is equipped with quick-ratio power steering, four-wheel independent suspension, and four-wheel disc brakes.

All this power, all this technology is yours in one superbly designed sports coupe. At your fingertips, you have the power to command one of the most responsive cars on the road today — Eagle Talon TSi AWD. For further information on Eagle Talon, call 1-800-JEEP-EAGLE.

At Your Eagle Dealer.

7/70 Protects engine and powertrain for 7 years or 70,000 miles and against outerbody rust-through for 7 years or 100,000 miles. See limited warranty at dealer. Restrictions apply.

The Mammalian Choroid Plexus

It serves as a "kidney" for the brain, bathing the delicate cells in chemically stable fluid. Although the choroid plexus is small, its role in nourishing and protecting the nervous system is great

by Reynold Spector and Conrad E. Johanson

The brain is choosy about what it allows to enter its tissues. Because it can function only in a chemically stable environment, the brain shuts out most water-soluble substances in the blood—even some potentially helpful medications. Three structures stand as gatekeepers to the brain's interior: the network of cerebral capillaries, the arachnoid membrane that covers the brain's surface and the diffuse, highly vascularized tissue called the choroid plexus.

The choroid plexus and the arachnoid membrane act together as barriers between the blood and the cerebrospinal fluid (CSF), a watery broth of nutrients, ions and other essential molecules. The CSF bathes the exterior of the central nervous system and fills the ventricles, the four large cavities inside the brain. Because the CSF exchanges substances freely with the interstitial fluid that surrounds the brain's neurons and supportive glial cells, the blood-CSF barrier is vital for keeping dangerous compounds out of the brain. The arachnoid membrane is generally impermeable to water-soluble substances, and its role in forming a blood-CSF barrier is largely passive.

REYNOLD SPECTOR and CONRAD E. JOHANSON have made major contributions to the study of the choroid plexus and the blood-CSF barrier. Spector is executive director of clinical sciences at Merck Sharp & Dohme Research Laboratories. He received his medical degree from Yale University in 1966 and has devoted much of his career to studying how the choroid plexus regulates the concentrations of various molecules within the nervous system. Johanson, who received his Ph.D. in physiology from the University of Kansas Medical School, is professor of clinical neurosciences at Brown University. He is also director of the Cerebrospinal Fluid Research Laboratory at Rhode Island Hospital in Providence, where he studies neurohormonal influences on ion transport in the choroid plexus.

The choroid plexus, however, actively regulates concentrations of molecules in the CSF and makes the blood-CSF barrier a selective one.

In effect, the choroid plexus is like a "kidney" for the brain: it maintains the chemical stability of the CSF in much the same way that the kidneys maintain the stability of the blood. The choroid plexus is not simply an excretory organ, however. It also manufactures CSF and provides the fluid with nutrients extracted from the blood. The behavior of the choroid plexus is quite different from that of the cerebral capillaries, which constitute the blood-brain barrier and mediate the diffusion of substances from the blood directly into the interstitial fluid [see "The Blood-Brain Barrier," by Gary W. Goldstein and A. Lorris Betz; SCIENTIFIC AMERICAN, September, 1986]. Nonetheless, the functional goals of the choroid plexus and the cerebral capillaries are the same: to provide the brain with a stable, nourishing milieu.

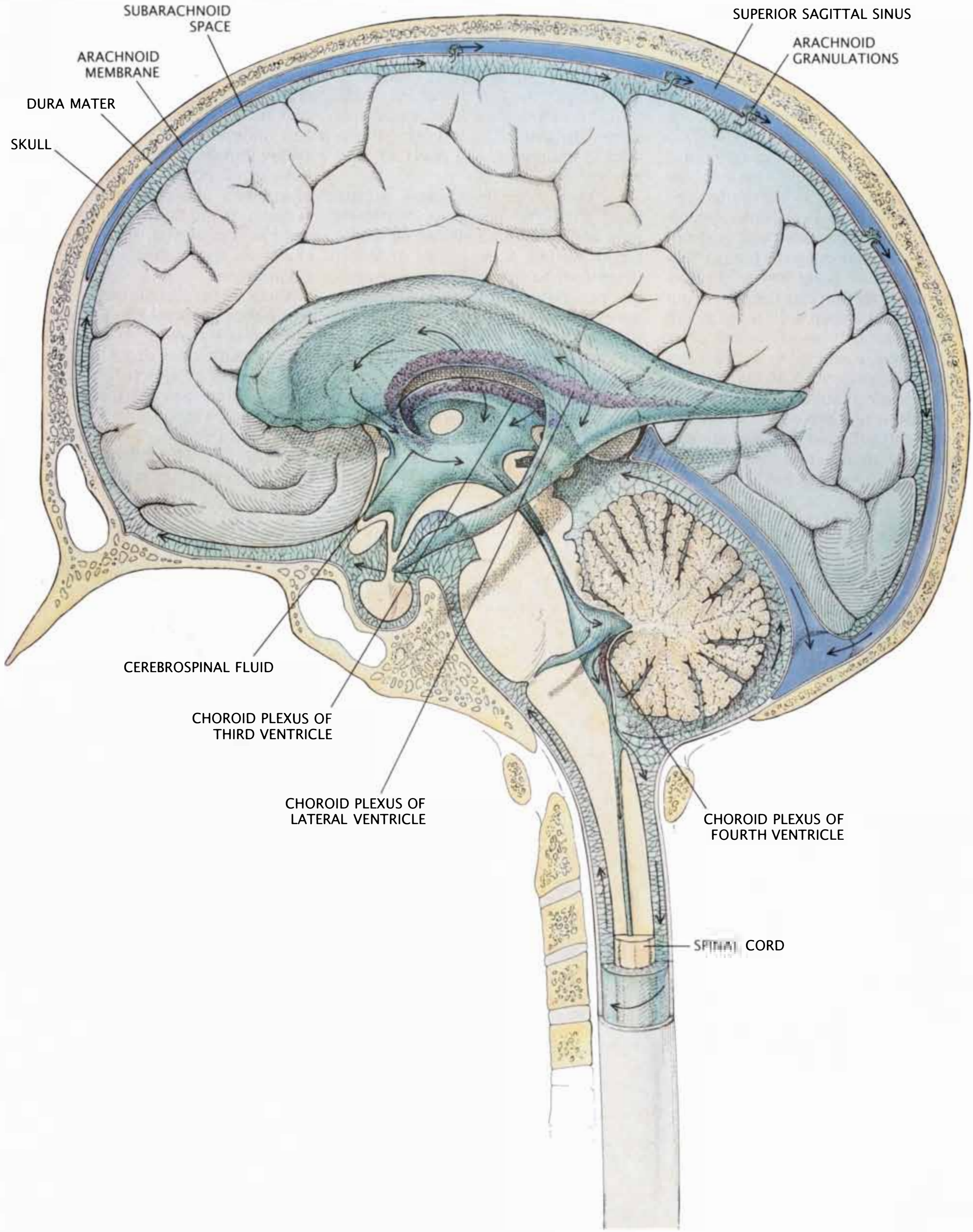
We and our colleagues have studied the physiology of the choroid plexus, including the regulatory mechanisms that choreograph the movements of substances through it and the differences between the blood-CSF barrier and the blood-brain barrier. As the workings of the choroid plexus become clearer, it should be possible to improve therapies for central nervous system disorders and for infections of brain tissue, including acquired immunodeficiency syndrome (AIDS).

In human beings and other mammals, the choroid plexus consists of several small, reddish tufts or patches of tissue. Most of the choroid plexus is almost equally distributed throughout the fourth ventricle near the base of the brain and the lateral ventricles inside the right and left cerebral hemispheres; about a tenth of the choroid plexus is in the centrally located third ventricle. In most adult mammals, the choroid plexus weighs only about .25 percent as much as the entire brain; in human beings, whose brains weigh slightly more than one kilogram, the choroid plexus adds up to two or three grams.

Viewed under a light microscope, the choroid plexus appears as a network of densely branched fronds. Each frond consists of capillaries and other small blood vessels surrounded by a single layer of epithelial cells. Like the epithelial cells that line the intestines and other organs, those of the choroid plexus are structurally polarized. One side of each cell, called the basolateral surface, is in contact with blood plasma that filters through the "leaky" walls of the choroid plexus capillaries. The opposite side of the cell, called the apical surface, puckers into villi, or microscopic fingerlike projections, that extend into the ventricular CSF.

Adjacent choroid epithelial cells are sealed together by intercellular structures called tight junctions that impede the passage of even small water-soluble molecules between the blood and the CSF. Tight junctions are the anatomical foundation of the blood-CSF barrier. They serve a similar function between the endothelial cells lining the cerebral capillaries, where they create the blood-brain barrier.

The general impermeability of the blood-CSF and blood-brain barriers makes it possible for essential biochemical functions in the brain to be compartmentalized and closely regulated. Some ions and nutrients are purposefully allowed through the barriers, and these diffuse slowly through the CSF and the interstitial fluid. These two extracellular fluids tend to have similar compositions because they are separated by only a single layer of cells, called ependymal cells, lining the ventricles. Unlike the tightly sealed layer of choroid epithelial cells, these ependymal cells are loosely linked by gap junctions, which are porous



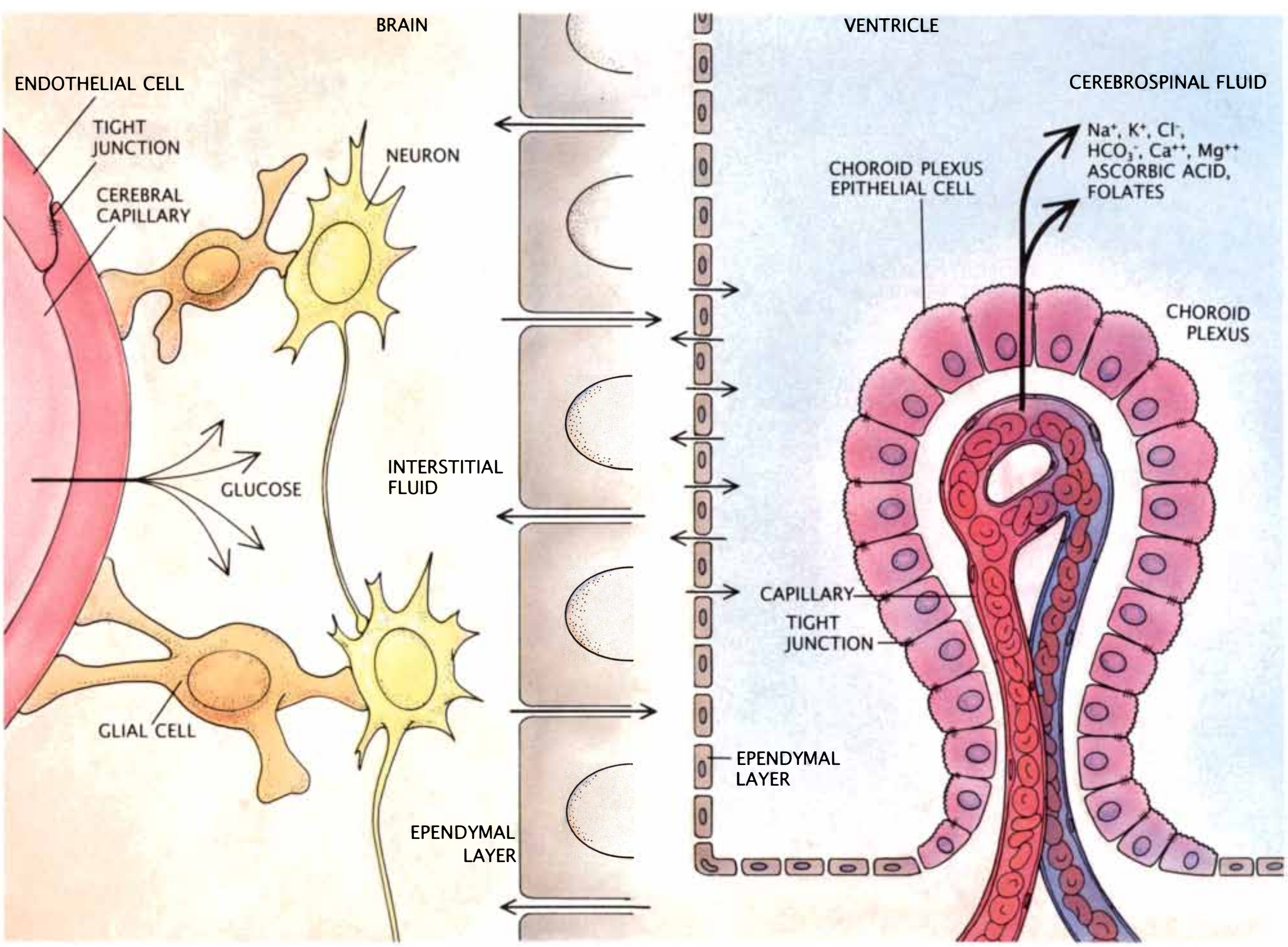
CHOROID PLEXUS sits within the fluid-filled chambers in the brain called ventricles. It continuously secretes cerebrospinal fluid (CSF), which cushions the nervous system, carries some nutrients to brain tissues and cleanses the brain of wastes. As new CSF is produced, old CSF is forced to flow (*arrows*) through the ventricles, around the spinal cord and into the subarachnoid space around the brain. As the CSF flows, it exchanges substances with the interstitial fluid between brain cells. Eventually it drains into blood in the superior sagittal sinus through structures called arachnoid granulations. Because the choroid plexus and the arachnoid membrane stand between the blood and the CSF, they constitute the blood-CSF barrier.

and negligibly impede diffusion. This "leaky" ependymal layer thus permits intimate contact between the interstitial fluid in the brain tissues and the CSF in the ventricles.

The choroid plexus performs a number of essential functions. Foremost among these is the secretion of CSF. As already noted, the CSF provides brain cells with a specialized, chemically stable fluid environment. It also buoys up the brain, effectively reducing brain weight 30-fold. This buoyancy protects the brain from injuries that would otherwise result from abrupt movements of the head. The choroid epithelial cells manufacture up to 90 percent of the CSF; the rest is made elsewhere in the brain. Every gram of choroid epithelial tissue secretes fluid at a rate of approximately .4 milliliter each minute. In adult human beings, this is sufficient to replace the total volume of the CSF (about 150 milliliters) every three or four hours. As new fluid continuously forms, the old CSF is displaced to the top of the brain, where it drains into the venous blood through membrane structures, such as the arachnoid granulations, which act as one-way valves.

To make CSF, the choroid plexus must draw the necessary nutrients, ions and other ingredients from the blood plasma. A brisk flow of blood (from four to five milliliters per minute per gram of tissue) is therefore necessary to sustain the required CSF output. Blood flow to the choroid plexus can be substantially changed by many drugs, such as vasopressin, and by various neurotransmitters, such as norepinephrine. In 1985 Vincent A. Murphy, now at the Laboratory of Neurosciences at the National Institutes of Health, and one of us (Johanson) studied the sensitivity of the blood-CSF barrier to changes in blood pressure. We were able to show, by injecting norepinephrine into rats, that a rapid increase in arterial blood pressure past a threshold of from 160 to 170 millimeters of mercury disrupted the blood-CSF barrier blocking the transfer of albumin. Presumably, fluid pressure above the threshold level distends the tight-junction seals in the choroid plexus, thereby causing chemical instabilities in the brain's environment.

About 99 percent of the CSF is water, which the choroid plexus secretes into the ventricles by creating ion gradients on both sides of its membranes. The process by which this occurs is complex, but it is possible to offer a simplified version. Water molecules in the choroid epithelial cells dissociate into hydrogen (H^+) and hydroxyl (OH^-) ions. The hydroxyl ions combine with



ESSENTIAL NUTRIENTS reach the neurons and glial cells in the brain by crossing either the blood-CSF barrier, which is regulated by the choroid plexus, or the blood-brain barrier of the cerebral capillaries. Water-soluble molecules cannot diffuse freely between the blood and the CSF because of impermeable tight-junction seals between the choroid epithelial cells; instead the epithelial cells transfer certain molecules from one side of the barrier to the other. Once molecules enter the CSF, they can diffuse through the "leaky" ependymal layer and reach the interstitial fluid around the neurons and glial cells. Similarly, wastes in the interstitial fluid can pass into the CSF for disposal. The endothelial cells of the cerebral capillaries, which are also sealed by tight junctions, control the direct exchange of materials between the blood and interstitial fluid.

intracellular carbon dioxide (a product of cell metabolism) to form bicarbonate (HCO_3^-) ions. At the basolateral surface of the cells, the hydrogen ions are then swapped for extracellular sodium (Na^+) ions from the blood plasma. The captured sodium ions are then pumped out the opposite side of the cells, through the apical surface, into the ventricles. This flux of sodium ions creates an excess of positive charge in the ventricles; negatively charged chloride (Cl^-) and bicarbonate ions move into the ventricular space to neutralize it. To maintain osmotic balance, water then diffuses into the ventricles as well. The final result of these activities is that water and many of the ions needed by the brain are secreted into the ventricles as CSF.

Concentrations of sodium, potassium, calcium, magnesium and chloride ions in the CSF do not change much even under extreme dietary and environmental conditions. As the CSF flows through the ventricles, its composition changes minimally because of diffusion through the ependymal layer. More important, the precise concentrations of ions in newly formed CSF stay roughly the same even when concentrations in the blood are fluctuating; this can occur only if mechanisms in the choroid plexus are adjusting the rates of ion transport between the blood and the CSF.

In a series of studies, one of us (Johanson) and Quentin R. Smith, now at the Laboratory of Neurosciences at the National Institutes of Health, found further proof that ion transport is actively regulated by the choroid plexus. We measured the simultaneous concentrations of ions in samples of CSF and plasma from rats and determined the concentrations of these ions in the epithelium of the choroid plexus. Analysis of the ion gradients across the epithelial-cell membranes, together with findings from other experiments, revealed that something more than passive diffusion was at work: active mechanisms in the choroid plexus were establishing higher concentrations of sodium and chlorine ions in the CSF than in the plasma. More specifically, active mechanisms were rapidly moving chlorine and sodium ions into the epithelial cells across their basolateral membranes. Other specialized mechanisms in the apical membranes were facilitating the transfer of these ions into the CSF.

Ions and water are not the only substances that brain cells must obtain from the blood to remain healthy. Without adequate supplies of glucose for energy, the brain of a mammal will lose consciousness within minutes and die. Deficiencies of vitamins such as thiamine have well-documented neurological effects, including impairment of memory and paralysis of eye movements. Other molecules that are vital to the brain include amino acids, which are used to make proteins, and ribonucleosides and deoxyribonucleosides, which are the building blocks of RNA and DNA, respectively. Given the effectiveness of the blood-CSF and blood-brain barriers in keeping out water-soluble molecules, how can nutrients reach the brain cells?

The answer is that specific transport systems at the barriers move nutrients from the blood into the CSF or directly into the brain's interstitial fluid. The blood-brain barrier and the blood-CSF barrier do not work in the same way, however; each is specialized to carry specific categories of nutrients into the brain by different mechanisms.

The blood-brain barrier of the cerebral capillaries is principally in charge of transporting those substances that the brain consumes rapidly and in large quantities—glucose, amino acids, lactate and ribonucleosides. The endothelial cells making up the cerebral capillaries have distinct *facilitated-diffusion* systems for moving molecules into the brain quickly and directly. Facilitated diffusion is a process involving special "carriers" that transport molecules across a membrane without consuming energy. These carriers can be imagined as specialized protein "doors" that open selectively, allowing certain molecules to pass through freely while locking out others. Molecules flow in both directions across the membrane, but the net flow is from the side of higher concentration to that of lower concentration. Because the brain consumes glucose rapidly, the concentration of glucose in the interstitial fluid and in the CSF is usually lower than in the blood plasma. Hence, the net flow of glucose across the blood-brain barrier is generally in the desired direction.

In contrast, the choroid plexus is able to transfer different nutrients into the CSF by a more complex, circuitous method. In general, the choroid plexus controls the transfer of some "micronutrients": substances that are essential to the brain but only needed in relatively small amounts over extended periods. These include vitamin C, folates (members of the vitamin B-complex class), deoxyribonucleosides and vitamin B_6 (pyridoxine). (Small amounts of glucose and thiamine also cross the choroid plexus, probably by facilitated diffusion, but their contributions to metabolism in the brain are minimal.)

Concentrations of micronutrients in blood plasma are usually low; to scavenge them from the blood, the choroid plexus relies on *active-transport* mechanisms. Unlike facilitated diffusion, active transport is an energy-consuming process that pumps molecules "uphill" against their gradient, concentrating them on one side of the blood-CSF barrier. Pumplike carriers on the basolateral surface of choroid epithelial cells pull micronutrients out of the blood and into a cell's cytoplasm. Once the micronutrients are concentrated in the cytoplasm, they are released into the CSF through the cell's apical surface, probably by facilitated diffusion. The active-transport systems in the choroid plexus are efficient: those that carry vitamin C and folates across the cell membranes, for example, normally concentrate these nutrients in the CSF to levels four times higher than in the blood plasma.

Both types of mechanisms—facilitated diffusion in the cerebral capillaries and active transport in the choroid plexus—tightly regulate the concentrations of molecules moving into the CSF and the brain's interstitial fluid. Their regulatory abilities are a direct consequence of the fact that they rely on limited numbers of carriers to transport molecules.

In each system, if the blood-plasma concentration of a nutrient rises above a certain level, then all the carriers for the nutrient are occupied, and the transport mechanism is said to be "saturated." Further increases in the blood's nutrient concentrations cannot raise the concentrations in the CSF or the interstitial fluid significantly higher. In the choroid plexus, several of the transport systems for vitamins are half-saturated at normal vitamin concentrations in the blood; because even minor surges in the vitamin content of the blood will fully saturate these systems, concentrations of these vitamins in the CSF are sharply limited. Conversely, if the concentration of a micronutrient in the plasma becomes abnormally low, then the carriers in the choroid plexus pull proportionally more of the micronutrient molecules out of the blood to maintain the concentration in the CSF.

There are important differences between facilitated diffusion and active transport that bestow certain advantages on each mechanism. Because facilitated diffusion requires no added energy, it is ideal for transporting glucose and other nutrients that must be moved into the brain continuously and in large amounts. Active transport requires energy but delivers micronutrients to the CSF at higher concentrations than facilitated diffusion could. Furthermore, because active transport is strictly a one-way process, blood that is poor in micronutrients cannot drain them from the CSF as would happen if facilitated diffusion controlled their transfer. In effect, active transport and facilitated diffusion are complementary regulation systems, each uniquely contributing to the chemical stability of the brain.

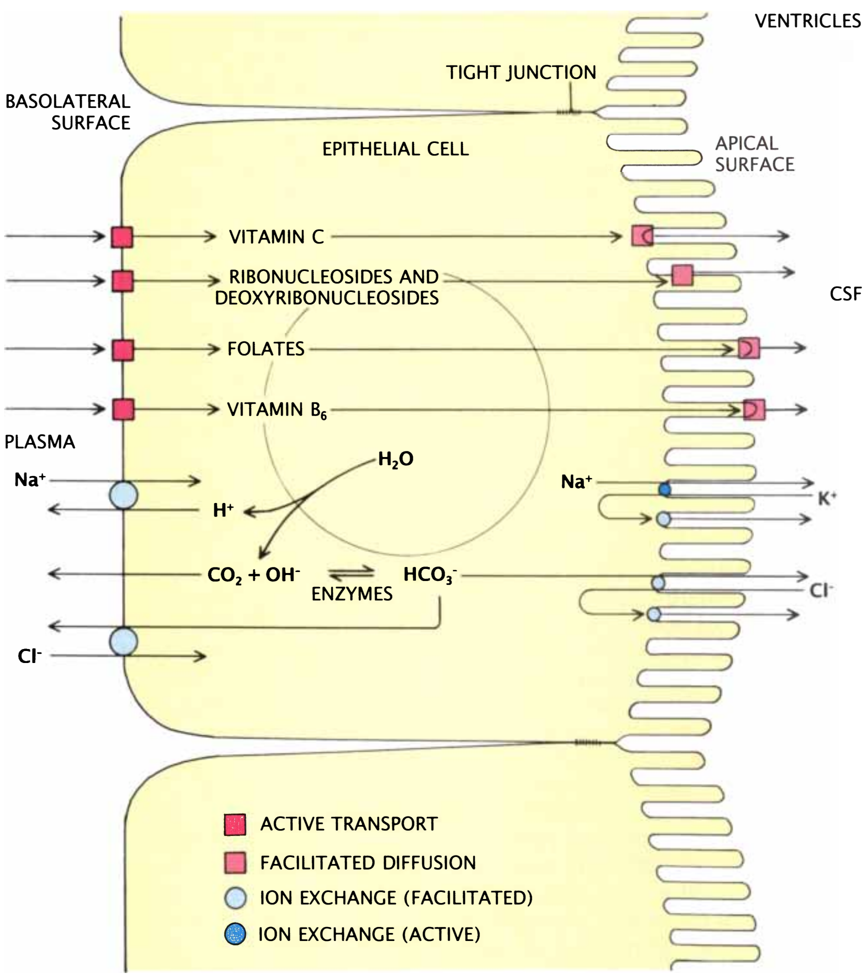
Why have the choroid plexus and the cerebral capillaries become so differently specialized? As yet there is no definitive answer, but we hypothesize that the route through the choroid plexus may have been more important for transporting glucose and other major nutrients at an earlier stage in mammalian evolution, when the forebrain was relatively small and the choroid plexus was proportionally large. There is evidence during fetal development, when the choroid plexus is huge and fills most of the lateral ventricles, that the choroid plexus may have an expanded role in nutrition. Fetal choroid epithelial cells contain significant amounts of glycogen, a molecule that may be converted into glucose and carried into the CSF. Yet with the great evolutionary expansion of the forebrain in human beings and other mammals, the choroid plexus may have become too small to provide adequate supplies of glucose and other nutrients to adult brain tissues. Thereafter, the larger network of cerebral capillaries was better able, it seems, to meet the major demands of the mammalian brain, whereas the choroid plexus was suited for the job of handling micronutrients.

To determine whether a micronutrient enters the brain primarily by way of active transport at the blood-CSF barrier, four criteria must be met. First, diffusion of the substance across the blood-brain barrier must be negligible. Second, when the substance is injected into the blood, it must rapidly accumulate in the choroid plexus and then in the CSF, followed by slower diffusion into the brain's interstitial fluid. Third, laboratory data on the specificity and behavior of the putative nutrient transport system in the choroid plexus must be consistent with the system's activity in an organism—that is, the transport system must be able to produce the nutrient concentrations observed in the living brain. Finally, the substance must be able to diffuse from the CSF to brain cells that may be separated from the CSF by tissue more than a centimeter thick.

One of us (Spector) and several colleagues at Harvard University and the University of Iowa have conducted experiments showing that all four of these criteria are met for vitamin C and that at least three of the four are met for the transport of deoxyribonucleosides, folates and vitamin B_6. The experiments also demonstrated important differences between various active-transport systems in the choroid plexus.

Vitamin-transport systems at the blood-CSF barrier, for example, are sometimes remarkably selective. The molecules of ascorbic acid (vitamin C) and isoascorbate (a common food preservative) are stereoisomers: they differ only in the geometric arrangements of their atoms. Yet according to Spector and Antonio V. Lorenzo of the Children's Hospital in Boston, if the blood holds equal concentrations of both molecules, the choroid plexus transfers 20 times more ascorbic acid than isoascorbate into the CSF.

In contrast, the transport of deoxyribonucleosides and ribonucleosides, which are different classes of molecules, is governed by a single system in the choroid plexus. (Ribonucleosides, but not deoxyribonucleosides, can also cross the blood-brain barrier.) Consequently, ribonucleosides and deoxyribonucleosides compete for access to the carrier molecules, and whichever nucleoside is most abundant in the blood plasma is transported into the CSF most frequently. Yet even this system displays some



FLOW OF MOLECULES across the blood-CSF barrier is regulated by several mechanisms in the choroid plexus. Some micronutrients, such as vitamin C, are pulled into the epithelial cells at the basolateral surface by an energy-consuming process known as active transport; the micronutrients are released into the CSF at the apical surface by another regulated process, facilitated diffusion, which requires no energy. Essential ions are also controllably exchanged between the CSF and blood plasma. Transport of an ion in one direction is linked to the transport of a different ion in the opposite direction, as in the exchange of sodium (Na^+) ions for potassium (K^+) ions.

selectivity: it will not transport cytarabine, a chemotherapeutic agent related to the nucleosides that are transported. It appears that the choroid plexus-transport system has an affinity only for nucleosides with certain structural characteristics.

Not all the tasks of the choroid plexus involve pumping materials from the blood into the CSF; sometimes it does just the opposite. The choroid plexus has an important role in cleansing the CSF of waste substances that form in the brain tissues as by-products of metabolic reactions. It is worth noting that although the choroid plexus and the kidneys do have analogous excretory functions, they move materials in opposite directions. The kidney removes waste products from the blood, whereas the choroid plexus pumps waste products into the blood.

The choroid plexus has several distinct active-transport "cleansing" systems. One system specifically removes iodide and other small inorganic ions from the CSF. Others clear the CSF of drugs including antibiotics such as penicillins and cephalosporins. Many compounds made within the brain, such as the metabolic derivatives of neurotransmitters, are also removed by the choroid plexus. Spector and Edward J. Goetzl of the University of California at San Francisco have found evidence that the choroid plexus transports leukotriene *C*4, an immunological compound involved in inflammation, out of the CSF and into the blood. If leukotriene *C*4 were to accumulate inside the brain, it could cause the brain to swell with excess fluid, a dangerous condition. Rapid transfer of this substance into the blood dissipates it and allows the liver or kidneys to detoxify it.

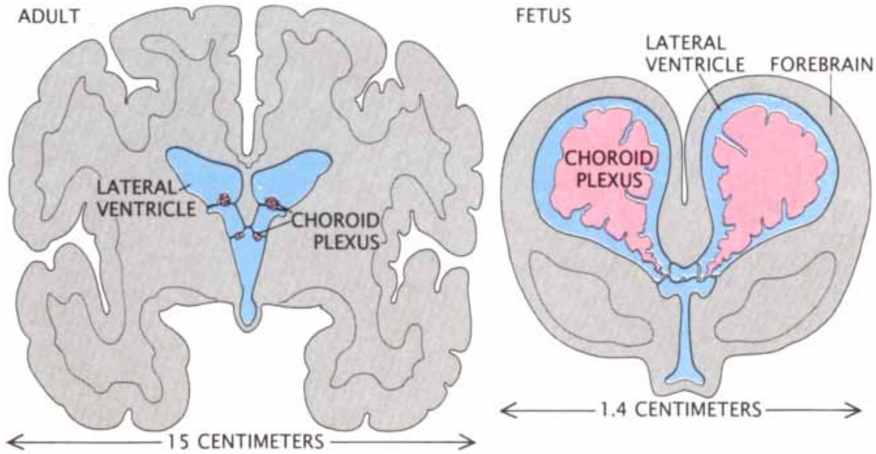
Some waste materials are literally rinsed from the brain by the steady flow of CSF from the choroid plexus. Because newly synthesized CSF contains little if any of these substances, there is a net diffusion of water-soluble waste molecules from the interstitial fluid into the CSF. Eventually the CSF, slightly contaminated by wastes, departs the central nervous system through the arachnoid granulations or other sites of drainage into the blood.

Another apparent function of the choroid plexus is the secretion of certain proteins. The overall concentration of protein in the CSF is approximately .5 percent of that in the plasma because of the blood-CSF and blood-brain barriers; however, several proteins are present in much higher concentrations in the CSF. One such protein is prealbumin, which is thought to help transfer thyroid hormones and vitamin A from the blood to the brain. Gram for gram, the choroid plexus contains at least 100 times as much prealbumin messenger RNA as the liver. No prealbumin messenger RNA is found in other parts of the brain. Because prealbumin messenger RNA is a genetic precursor of prealbumin protein, these results suggest that the choroid plexus synthesizes prealbumin and then transfers it to the CSF.

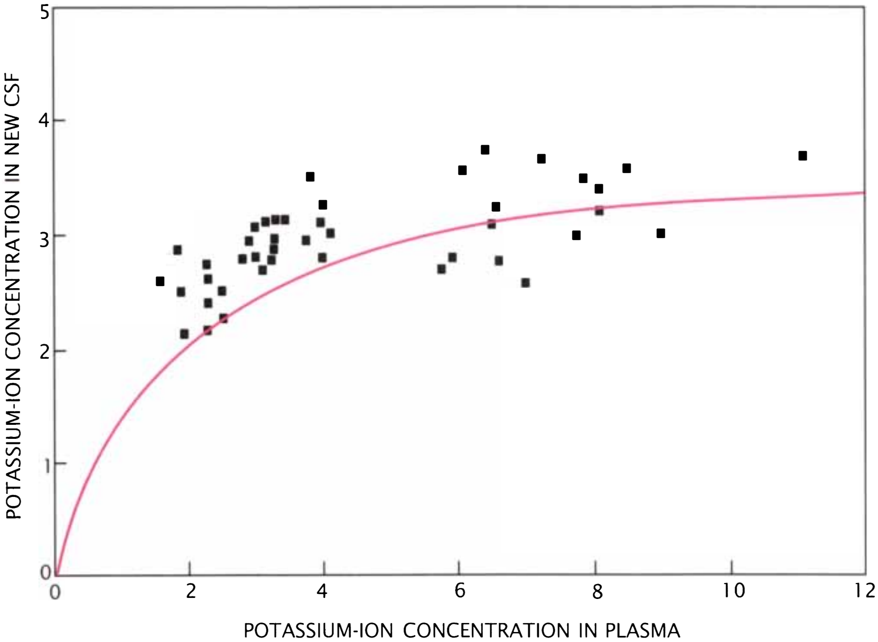
Other proteins secreted by the choroid plexus may play important roles in some aspects of brain development. Three decades ago the Russian physiologists N. S. Volzhina and Boris N. Klosovsky observed that the nervous system of a puppy atrophied if the choroid plexus was removed. The effects of proteins released by the choroid plexus is an exciting and potentially rich area in neurobiology for future investigation.

Given the many ways that the choroid plexus helps to maintain the health of the brain, it is not surprising that disruptions of this tissue's functions can cause serious illness. Large tumors in the choroid plexus may obstruct the flow of CSF or compromise cerebral circulation by pressing against brain tissue; one rare tumor called a benign adenoma manufactures too much CSF. Such adenomas may secrete from three to five times as much CSF as normal, overwhelming the brain's mechanisms for reabsorbing the fluid. The result is hydrocephalus, more commonly known as "water on the brain." As the extra CSF accumulates in the ventricles, it puts abnormally high pressure on the brain tissues and damages them. Although there are drugs and fluid-shunting procedures for treating this condition, removal of the hypersecreting adenoma is the best solution if surgically possible.

The choroid plexus plays a dual role in the treatment of patients with bacterial meningitis, at some times helping and at other times hindering their cure. In this type of meningitis, bacteria infiltrate the CSF and the choroid plexus, where they cause inflammation. As previously noted, penicillins and most cephalosporin antibiotics are actively removed to varying degrees from the CSF; the concentration of penicillin G in a healthy mammal's CSF is only about 1 percent of that in the plasma. Ordinarily, then, the blood-CSF barrier makes it difficult for penicillin G to kill bacteria in the brain. Spector and Lorenzo have shown, however, that inflammation during meningitis causes the choroid plexus to develop "leaks," which leads to a partial breakdown of the blood-CSF barrier. As a result, concentrations of penicillin G in the CSF of mammals with meningitis may vary between 2 and more than 20 percent of those concentrations in the plasma. In many cases, such concentrations of penicillin are sufficient to kill or retard the growth of bacteria in the CSF. But as the inflammation subsides and the choroid plexus becomes fully functional again, it resumes vigorously pumping penicillin out of the CSF. If all the bacteria in the CSF are not dead when choroid plexus function returns to normal, a relapse of the infection is possible. For this reason, physicians prefer to treat



FETAL CHOROID PLEXUS is proportionately larger than that of an adult human being and fills more of the space in the ventricles. This observation suggests that the choroid plexus may have a particularly large role in nourishing the brain during development; it may also have been more important at earlier stages of evolution.



CONCENTRATION of potassium (K^+) ions in freshly secreted CSF does not change by much even when the concentration in the blood rises and falls dramatically. Potassium-transport mechanisms in the choroid plexus are nearly saturated at normal concentrations in the blood: almost all the carriers are busy transporting potassium ions. Increasing the level of potassium in the blood has little effect on the level in the CSF because there are few spare carriers to move the additional ions. These data points were collected by Adelbert Ames III of Harvard Medical School.

many types of meningitis with antibiotics such as chloramphenicol that typically reach high levels in the CSF or with antibiotics such as ceftriaxone that are not actively transported out of the CSF as readily as penicillin G.

A more speculative effect of pathological changes in the choroid plexus, namely, impaired cerebral functioning, may occur in the elderly. There is substantial evidence that the choroid plexus, like many other organs, deteriorates with advanced age. For that reason, some investigators suspect that certain types of degenerative brain disease may be caused in part by inadequate choroid plexus function and accompanying cerebral malnutrition. The idea is still greatly in need of experimental verification, however.

As the role of the choroid plexus in sickness and health becomes better understood, it is apparent that manipulation of its various functions holds great promise for clinical applications. The growing body of knowledge about the effects of neurotransmitters and hormones on the activity of the choroid plexus can guide the development of new treatments.

Altering the rate at which CSF is secreted is an area worthy of further research. Most attempts to alter CSF production have sought to inhibit its flow, in large part because drugs that exert this effect have obvious applications in treating hydrocephalus and cerebral edema, conditions in which accumulating fluid damages the brain.

Manipulating the balances of various ions in the CSF could be similarly beneficial. It is known, for example, that when the concentration of potassium ions in the CSF is high, neurons are more easily excited and fire more readily. Some anticonvulsant drugs for epilepsy, such as phenytoin, may work at least in part by increasing the transport of potassium ions out of the CSF through the choroid plexus. Calcium and magnesium ions also affect rates of neuron firing; in addition, they induce the dilation of cerebral arteries. If more can be learned about the flow of these ions through the blood-CSF barrier, then perhaps someday drugs will be devised that can shift the concentrations of calcium or magnesium ions to minimize the risks or consequences of epilepsy.

The treatment of AIDS should also improve with a fuller understanding of transport mechanisms and their regulation in the choroid plexus. It is now clearly established that the virus causing AIDS can invade the central nervous system; one common consequence of this invasion is dementia. By hiding inside the brain, the AIDS virus finds sanctuary from many antiviral agents that cannot pass either the blood-CSF or the blood-brain barrier. Any prospective therapy for AIDS is doomed to fail, then, unless it can reach brain cells infected by the virus.

Fortunately, one of the most effective agents that is currently prescribed for AIDS—azidothymidine (AZT), also known as zidovudine—does cross the blood-CSF barrier and enter the brain. AZT is a nucleoside, and it appears to have an affinity for the nucleoside-transport system in the choroid plexus. To date, it is the first antiviral agent known to suppress the AIDS virus in human brain tissue. Still, the effectiveness of AZT and other drugs against AIDS might improve if they could be made to reach the brain's interstitial fluids faster and in greater concentrations. It is our hope that further research on the choroid plexus and the blood-CSF barrier will uncover techniques for enhancing the transport of AZT into the central nervous system or, alternatively, may encourage the development of new nucleoside agents that penetrate the barriers more readily.

When the physiology of the choroid plexus was less understood, the tissue often looked small and passive when compared with the brain's more extensive network of cerebral capillaries. Today it is clear that the choroid plexus is a major transfer point in the exchange of certain proteins, vitamins, ions, amino acids, nucleosides and other molecules between the blood and the CSF. Its role in preserving the stability of the central nervous system is unique. The clinical benefits that can spring from a fuller understanding of the choroid plexus and the blood-CSF barrier are profound.

FURTHER READING

THE CHOROID PLEXUS IN HEALTH AND DISEASE. Martin G. Netsky and S. Shuangshoti. University Press of Virginia, 1975.

VITAMIN HOMEOSTASIS IN THE CENTRAL NERVOUS SYSTEM. Reynold Spector in the *New England Journal of Medicine*, Vol. 296, No. 24, pages 1393-1398; June 16, 1977.

POTENTIAL FOR PHARMACOLOGIC MANIPULATION OF THE BLOOD-CEREBROSPINAL FLUID BARRIER. Conrad E. Johanson in *Implications of the Blood-Brain Barrier and Its Manipulation*, Vol. 1. Edited by Edward A. Neuwelt. Plenum Press, 1989.

MICRONUTRIENT HOMEOSTASIS IN MAMMALIAN BRAIN AND CEREBROSPINAL FLUID. Reynold Spector in *Journal of Neurochemistry*, in press.


DESKTOP PUBLISHING
NEWS
Open
Hou
TRAVELER

SCIENCE AND BUSINESS

Profitable pollution, flying neural nets, selling space, freer markets

Polluting Rights

How to build a market for an unwanted commodity

On the banks of the Ohio River sits the General James M. Gavin utility plant. It is the largest plant in Ohio, consuming about 5.6 million tons of coal every year and churning out more than 12.5 billion kilowatt-hours of electricity. Gavin has another distinction as well: it is the largest U.S. emitter of sulfur dioxide, a key contributor to acid rain. As such, it is a linchpin in the government's proposed plan to combat acid rain.

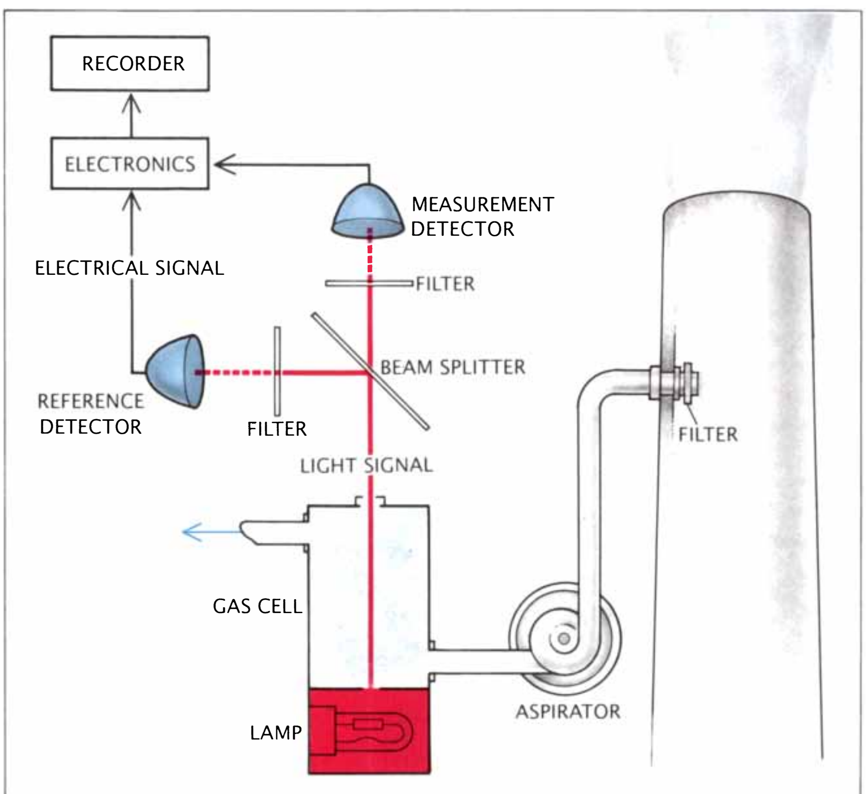
In July, President Bush unveiled a plan to reduce the country's annual sulfur dioxide emissions by 10 million tons (from a 1980 baseline of 25 million tons annually) by the end of the year 2000. Congress seems reconciled to the goal, which, says William K. Reilly, administrator of the Environmental Protection Agency, should "reverse the deterioration of our lakes and streams" caused by acid rain.

To reach that goal, the administration has turned to a policy long advocated by economists: grant "pollution allowances" that let the holders emit a specific amount of a pollutant. The companies may then buy, sell and lease these credits. In theory, such an approach should satisfy pollution requirements at the lowest cost. Yet even supporters warn that the success of any plan will hang on precisely how it is implemented.

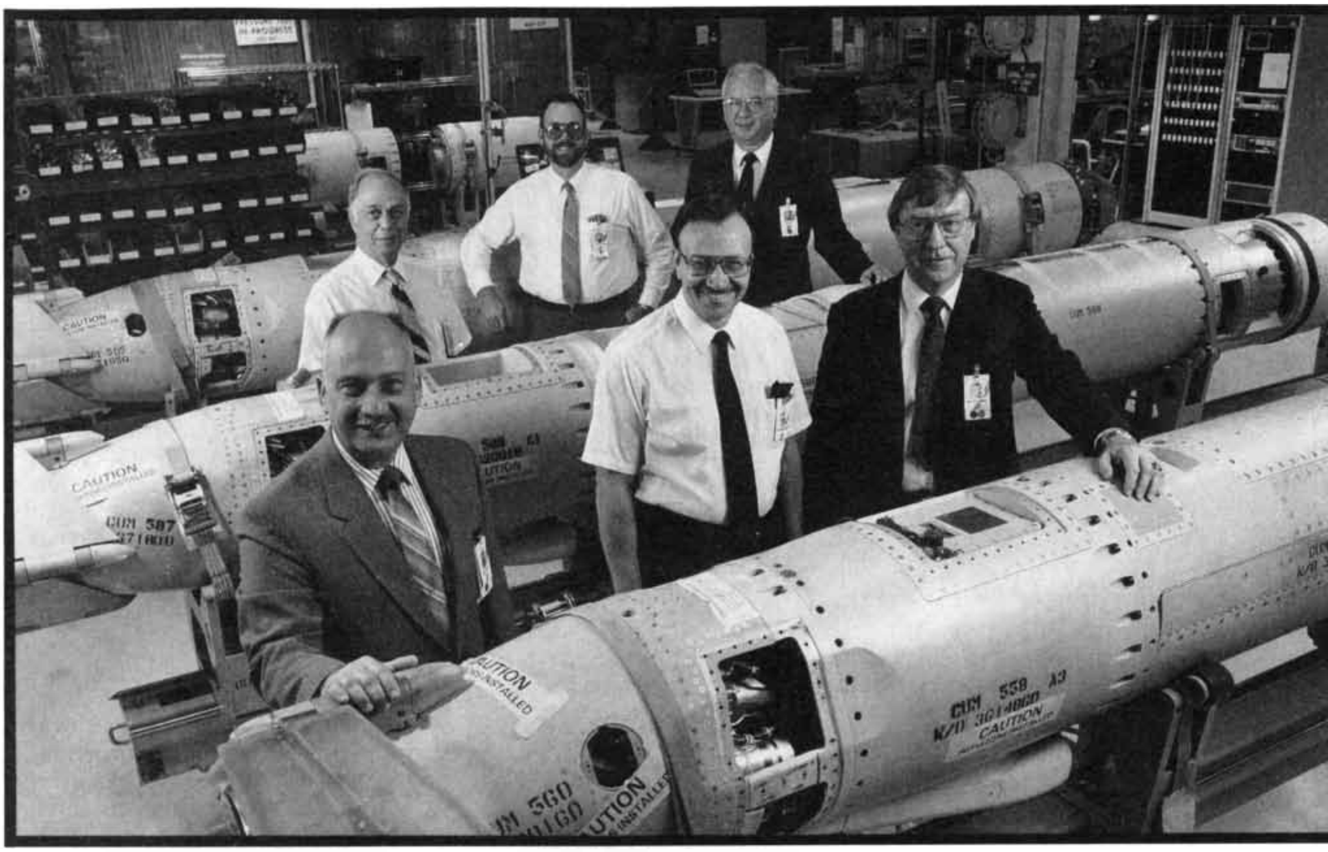
So far, the EPA has presented only a broad outline of the scheme. The main participants in the trading program will be coal-fired electric-power companies, which produce about two thirds of all U.S. sulfur dioxide emissions. (A diverse collection of industrial facilities accounts for the other third.) The EPA will initially give annual pollution allowances only to the largest and dirtiest facilities—about 107 plants in 20 states, including the Gavin operation. (The remaining 600 or so plants must still comply with existing standards.) Because each plant is sure to have too few allowances to cover its current level of emissions, it will be forced to cut emissions or buy allowances from other plants that are willing to make even deeper cuts. Since there is a fixed number of vouchers, even the dirtiest plants will eventually find it too costly to buy additional emissions allowances.

The first deadline these utilities face for making emissions match their vouchers is January, 1996. In this first phase, vouchers will limit plants to emitting the equivalent of 2.5 pounds of sulfur dioxide per million British thermal units of heat generated by the plants in 1985. By January, 2001, the EPA will have reduced its yearly issue of vouchers by limiting emissions to 1.2 lbs/million BTU and will have added more utilities to the program. At that point, the total annual utility emissions should be down to roughly nine million tons. The EPA says it will not let that figure creep up, even as the demand for electricity rises. New plants will have to buy allowances from existing facilities before they can emit any sulfur dioxide at all. And low-level polluters will not be permitted to increase the rate at which they emit sulfur dioxide.

A utility will have to go through an extensive review to get into the trading program. Once accepted, however, it will not need the EPA's permission to buy or sell allowances. (Until 2001, plants will be allowed to trade only with those in the same state or system; after that date they will be allowed to trade with any plant in the emissions-reduction program.) So, if Gavin agreed to sell or lease 500 allow-



***CONTINUOUS EMISSIONS MONITOR** measures the percentages of sulfur dioxide and nitrogen oxides in power-plant effluent. Gas is pumped from a smokestack into a chamber where ultraviolet or infrared light passes through it. The beam is split, and the light is directed to two detectors, one sensitive to a reference wavelength and the other to a wavelength that is absorbed by the pollutants. The ratio of the intensities of the two wavelengths indicates the concentration of pollutant.*



"Before we got working as a team, Rich and Marv would get together and agree that Terry had a problem. Terry and I would get together and agree that Joe had a problem. Once we all got together in the same room, we recognized that we *all* had a problem.

Production schedules for part of the Tomahawk missile's avionics system weren't what they should be. Assigned as a special troubleshooting team, it was our job to get the work flowing. That meant we had to talk with each other—plan together. And communicate—so one guy's solution didn't cause somebody else a new problem.

Within three months we began to see improvements. At six months, production really started to roll. And today we're producing two-and-a-half units in the time it had been taking to build just one. The key to our success was that we all accepted any one person's problem as everybody's problem."

—Howard Austin (front, left), Avionics Work Center Manager with Rich Heindl (front, center), Avionics Test Operations, Marv Prongue (front, right), DSMAC Engineer, Carl Hanks (rear, left), Section Manager Quality Engineering, Joe Porter (rear, center), Avionics Production Control, Paul Raidt (rear, right), DSMAC Program Manager, Terry Young (not present for photo), Manufacturing Engineering

MCDONNELL DOUGLAS

A company of leaders.

ances to its brother plant, Tanners Creek in Indiana, the two would work out terms and then send the EPA a letter announcing the transfer. The EPA would check a computer data base to ensure that Gavin had 500 credits to sell; it would then credit and debit the companies' accounts.

The trading program should spur utilities that can easily reduce emissions to make large cuts and sell the extra credits to companies that find it more expensive to cut their own emissions than to buy credits. The EPA consequently estimates that if companies traded allowances aggressively, the cost of the program—estimated at $4 billion—could drop by 20 percent.

As the EPA works out the details, it will be guided by powerful lessons from the handful of trading programs that already exist. California launched a plan in 1979 for emissions of smog-producing volatile organic compounds and other pollutants that has become a bureaucratic morass, observers say. Any new or expanded plant likely to emit significant levels of compounds must buy "offsets"—much like vouchers—from existing facilities. But local regulators must evaluate each trade, case by case. Documentation for the deals has run anywhere from 20 pages to several boxes of papers, says Malcolm C. Weiss, an attorney with McClintock, Weston in Los Angeles.

The EPA argues that its system will be far less cumbersome, since it does not plan to review each trade. Still, states may be wary that a local utility would buy credits—in compliance with national regulations—and pollute freely, hurting local conditions. Public utility commissions, which regulate the rates that generators charge for electricity, may also worry that plants are not using their allowances prudently. If local regulators enter the picture, the EPA's system is bound to become more complicated.

From an administrative vantage, a trading program in which different facilities are meeting different emissions requirements may create a complex monitoring task. To ensure that power stations do not emit more sulfur dioxide than their accounts allow, the EPA will install continuous emissions monitors on smokestacks. Although there are many varieties of CEM's, most shine infrared or ultraviolet light through effluent and measure the resulting decrease in the wavelengths of light absorbed by sulfur dioxide (and oxides of nitrogen, another contributor to acid rain).

But CEM's are not foolproof. Environmental engineers say much checking is needed to ensure that companies do not cheat. Enforcement is likely to be a job for local regulators, even though the EPA would keep track of each utility's overall "emissions account." The EPA and local regulators will therefore have to stay in close contact.

Whether utilities will make good use of their trading privileges depends on many factors outside the program's structure. In California, for instance, large companies often must be convinced that trading emissions credits is worth their while and will not subject them to closer regulatory scrutiny, says John Palmisano, a former EPA official, who now heads AER*X, a firm based in Washington, D.C., that brokers sales of emissions credits.

Companies may also be reluctant to trade credits if regulators change the rules. Although EPA officials in Washington claim they do not intend to change the national regulations, putting a guarantee into writing is exceedingly difficult, they say.

Utility companies, meanwhile, argue that the emissions cap will limit trading because it encourages companies to hoard allowances. Fearing shortages of credits for new or expanded power plants, companies will refuse to sell their credits, argues A. Joseph Dowd, general counsel for American Electric Power. The EPA counters that although there may be some squeeze on credits just after the turn of the decade, by 2010 enough aging plants will have been retired to ensure plenty of credits for new, cleaner facilities.

Legislators are also grappling with a host of broader questions: Should the program be extended beyond the utility industry? Should there be special provisions to protect the jobs of high-sulfur coal miners? Should provisions for reducing nitrogen oxides be stronger? Should companies be given special treatment if they invest in clean-coal technologies?

The very notion of pollution-rights trading worries some noneconomists. Representative Jim Cooper of Tennessee believes that the government's giving companies "permission" to pollute is tantamount to "trafficking in a morally objectional commodity." Nevertheless, many legislators, including Cooper, seem willing to give the idea a try. "The best push the government can give [to promote trading]," says Robert W. Hahn, a senior staff economist at the Council of Economic Advisers, "is to define the general ground rules, promote flexibility and then let the marketplace do what it will." —*Elizabeth Corcoran*

Nets Work

Fast, smart neural architectures start to show their prowess

Neural-network computing, in which multiple processors (real or simulated) modify one another's output, is far from a mature field. The approach, which is inspired by the architecture of living nervous systems, suffers from confused terminology and arguments over the capabilities of competing architectures. Still, companies bringing neural networks to bear on real-life problems are springing up. "People may be making money on [neural networks] before academics know it's possible," says Paul J. Werbos, a program officer at the National Science Foundation and a prominent neural-network theorist.

Unlike typical computers, neural networks can reach an answer by approximating or by generalizing based on limited data. As a result, they appear well suited to solving problems in pattern recognition or signal processing, which are often too murky for standard computing algorithms. Simple networks, for example, are already used to correct for signal distortions on varying telephone lines; they may eventually be used to pick up objects in random orientations from assembly lines. Now researchers are applying neural networks to tasks that exhibit nonlinear behavior over time—tasks in which a given action now can have complex effects later.

The more sophisticated systems approach such problems by combining at least two networks that employ a technique called backpropagation: the actual output of a network is compared with the desired output, and the discrepancy serves as a basis for modifying the interactions between "neurons" to improve the fit. The system in effect learns to adapt to the changing conditions.

Airline fare structuring provides an example of such a task: seats on a flight are shifted from one fare class to another as sales proceed in order to gain the best returns. Early bookings can be used to predict the demand for a flight and hence the optimum mix of cheaper and more expensive seats. Yet passengers in different fare classes tend to book at different times; moreover, flights at certain times, such as holidays, attract more early bookers looking for discount fares.

According to William R. Hutchison of BehavHeuristics in College Park, Md., a system based on several interacting neural networks, each with a different task, can outperform traditional statistical approaches to seat allocation by adapting continuously to changing consumer behavior. Nationair in Montreal already relies on BehavHeuristics' software package, Airline Marketing Tactician, for its seat allocations, and trials are under way with two major U.S. carriers. (Like most of today's networks, BehavHeuristics' system is actually a program that simulates a neural network on a general-purpose computer.)

Similar but faster-moving problems confront designers of autopilots. Neural Systems in Vancouver, B.C., has developed a stripped-down autopilot that is learning to control high-performance jets such as the F-16. To achieve a given aircraft response—level flight, say—the system compares the desired response with the plane's actual behavior and, through backpropagation, learns what adjustments to the controls are needed. When the network has learned how the aircraft responds, it can operate the controls directly and dispense with backpropagation. So far Neural Systems' autopilot has mastered only level flight and simple ascents and descents; turns will come later, says Gary M. Josin, the president of the company. (This may explain why the National Aeronautics and Space Administration, which supports Josin's work, has not yet let the network take the controls of a real plane.)

Other small companies are also finding aviation a promising area for networks. Accurate Automation Corporation based in Chattanooga, Tenn., is working on networks that could help air-traffic controllers to check that planes approaching an airport or an aircraft carrier are following the proper glide slopes. The company is also developing pattern-recognition systems to help identify aircraft radar signatures. (The identification of military targets is one of the most promising applications for neural networks, according to a study carried out last year by the Defense Advanced Research Projects Agency.)

Neural-network proponents say networks may prove a boon to robotics, where they could reduce the amount of computation needed for the "inverse dynamics problem": controlling the actuators in a robot arm so that the "hand" ends up at the right point. Although the problem is conceptually simple, the time conventional computers require for the calculations can limit robot performance and lead to inefficient, jerky motions. Josin has already shown how neural networks using backpropagation can quickly learn the needed geometric transformations from a few examples and can generalize that knowledge to fit new situations. Neural networks will, he predicts, eventually make robots "graceful." —*Tim Beardsley*

Space Available

U.S. industry edges into the space race

The U.S. commercial space program is flying, at long last.

In 1983 the U.S. government promised to encourage the development of a private-sector industry for launching payloads. The first such flight finally took place in late August, when McDonnell Douglas lofted a spacecraft for the British Satellite Broadcasting consortium into geostationary orbit. According to the jubilant Office of Commercial Space Transportation (part of the Department of Transportation), there may be as many as five launches of commercial rockets this year and almost two dozen launches next year.

Now begins the tough part. Even though space entrepreneurs have high hopes of a large market, they concede it is unlikely to be big enough for all of them. And there are still hiccups as the Air Force and the National Aeronautics and Space Administration learn to work with the private sector.

More than half a dozen U.S. companies are already competing in the worldwide rocket market, which will probably top $1.5 billion next year. The European company Arianespace is likely to win more than half of that business. Three long-time U.S. aerospace firms—McDonnell Douglas, Martin Marietta and General Dynamics—are also vying for shares. Meanwhile a fleet of small, privately held companies are competing for payloads that weigh only a few hundred pounds.

At first, the government is likely to provide roughly two thirds of the payloads that fly on U.S. rockets. The glut of unlaunched government missions that have accumulated since the space shuttle *Challenger* blew up in 1986 will sustain the large carriers until 1992, predicts Courtney A. Stadd, an aerospace consultant in Washington, D.C. As that backlog diminishes, the large carriers may also face increased competition from Soviet and Chinese launchers.

"It's uncertain how long we're going to be able to stay in the business," concedes Louis C. Raburn, manager of the McDonnell Douglas Delta commercial launching program. McDonnell Douglas is now reviewing the cost of building a new generation of launchers but has no final plans.

The fledgling launching companies, which include Space Services, Inc. (SSI), the American Rocket Company (AMROC) and Orbital Sciences Corporation, are also counting on government business. But they hope that a larger market for their services will open up when satellite manufacturers develop lightweight, relatively cheap satellites that are well suited to ride on board the smaller boosters those companies offer. "It's only going to take another two or three years for these satellites to become available," predicts David W. Thompson, president of Orbital Sciences.

The Air Force and NASA are still learning to be sensitive to this newly active private sector. For instance, even though other launching companies rely on proved rockets originally developed for government programs, AMROC resurrected an unproved design for a hybrid rocket motor that combines solid- and liquid-fuel technology. The patent rights for the concept had expired; AMROC spent almost four years turning the idea into a workable design. To the company's dismay, NASA recently decided to pursue a similar hybrid propulsion project. Acting President James C. Bennett asks: "Why should we put our money in" if the government is going to pay others to do similar work?

"NASA didn't realize the impact of its contracts on AMROC," Stadd says. The agency subsequently rewrote its plans to avoid competing with AMROC. "It's part of the transition problem," says a source at the Transportation Department. "In the past, NASA never had to answer to anyone."

"We're still operating in a bureaucratic society," Raburn says. Government agencies, too, are hampered in their dealings with private launch firms by hundreds of regulations, developed over many years, space entrepreneurs acknowledge.

Still, much red tape has been eliminated since the Transportation Department established its commercial space office, points out Donald K. ("Deke") Slayton, president of SSI and a former astronaut. He remembers when SSI first applied for permission to launch a rocket in 1982: at that time the company had to secure permits from 18 government agencies, including an export license (for taking technology beyond U.S. borders) and a firearm license (for the rocket). —*E.C.*

THE ANALYTICAL ECONOMIST

Markets unbound?

Command economies are unworkable; only free markets can provide incentives for growth and development. That is today's conventional wisdom as Poland, Hungary and even the Soviet Union lurch toward private enterprise. It has been the rationale for many years as the International Monetary Fund along with the World Bank has required the governments of poor countries to liberate their markets as a condition for continuing aid. (Completely unfettered economies have been tried and found wanting; in the U.S., rules curbing monopolies, insider trading, child labor and a host of other abuses testify to the need for some government intervention.)

The rejoicing over the trend toward free markets begs the question of just what a free market is. Economists' attitude toward the term has typically been that of late Supreme Court Justice Potter Stewart toward pornography: they know a free market when they see one. Only a handful of scholars have attempted a more rigorous definition, and "none of them cite each other," laments Frederic L. Pryor, an economist at the Hoover Institution at Stanford University.

The simplest measure of a free market, suggests one World Bank economist, is how much of a nation's output the government controls: the percentage of industrial production coming from government-owned plants and the percentage of agricultural production that flows through government marketing boards. As China has liberalized its economy, for example, the Bank's estimate for industrial control has dropped from about 97 percent to about 80 percent. Such numbers, however, do not tell the whole story: Thatcherite Britain, according to indexes published by conservative think tank Freedom House, is roughly as socialist as Nicaragua.

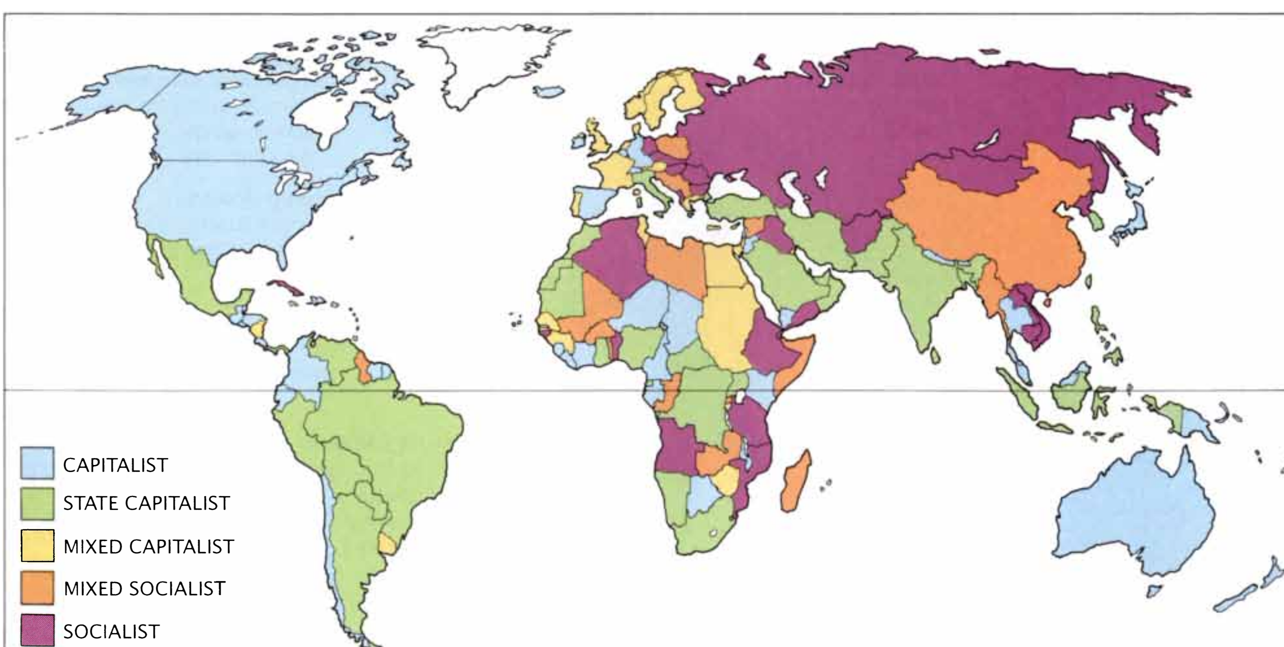
Governments can still control markets without owning them, points out political scientist Bruce M. Russett of Yale University. Typical restraints range from regulations to taxes to the occasional word to the wise. In the U.S. it used to be called jawboning; in Japan it is industrial policy.

Governments that control economic life may simultaneously control other activities, but correlations between economic and political freedom are not clear. The Soviet Union has increased liberty while maintaining a largely controlled economy; China apparently pursues free markets while preserving tight political control (although some Chinese companies and entrepreneurs are now being repressed). Governments on the right, Russett notes, may only appear to espouse free markets; they leave capital alone, but they interfere with labor markets by squelching unionists.

Some measures of economic freedom attempt to quantify the results of government interference: researchers may compare official and black-market prices, for example. The U.S. government estimates how many millions of hours companies spend filling out forms. Pryor starts out by tracking how much prices fluctuate with changes in demand. Until recently food prices in the U.S.S.R., for instance, remained constant from season to season and year to year.

Economists give many rationales for the superiority of free markets: personal incentives, freedom from administrative overhead, the impossibility of properly running a system as vast as an economy from a central point. A free market replaces the bureaucracy with the "hidden hand" of self-interest.

The empirical evidence, of course, varies. Whereas countries with rigidly controlled economies have generally done poorly, some ostensibly free-market ones have also fared ill. Active state guidance, meanwhile, has played a visible role in such miracles of capitalism as the economies of South Korea and Japan. —*Paul Wallich*



NATIONAL ECONOMIES are colored according to degree of government control in this map based on analyses compiled in 1988 by Freedom House, a conservative think tank. Systems include capitalist, mixed capitalist (government provision of social services), state capitalist (large government-controlled enterprises), mixed socialist (government control with a strong private sector) and socialist. Freedom House has prepared the most comprehensive list of this type.

SCIENCE IN PICTURES

Neptune

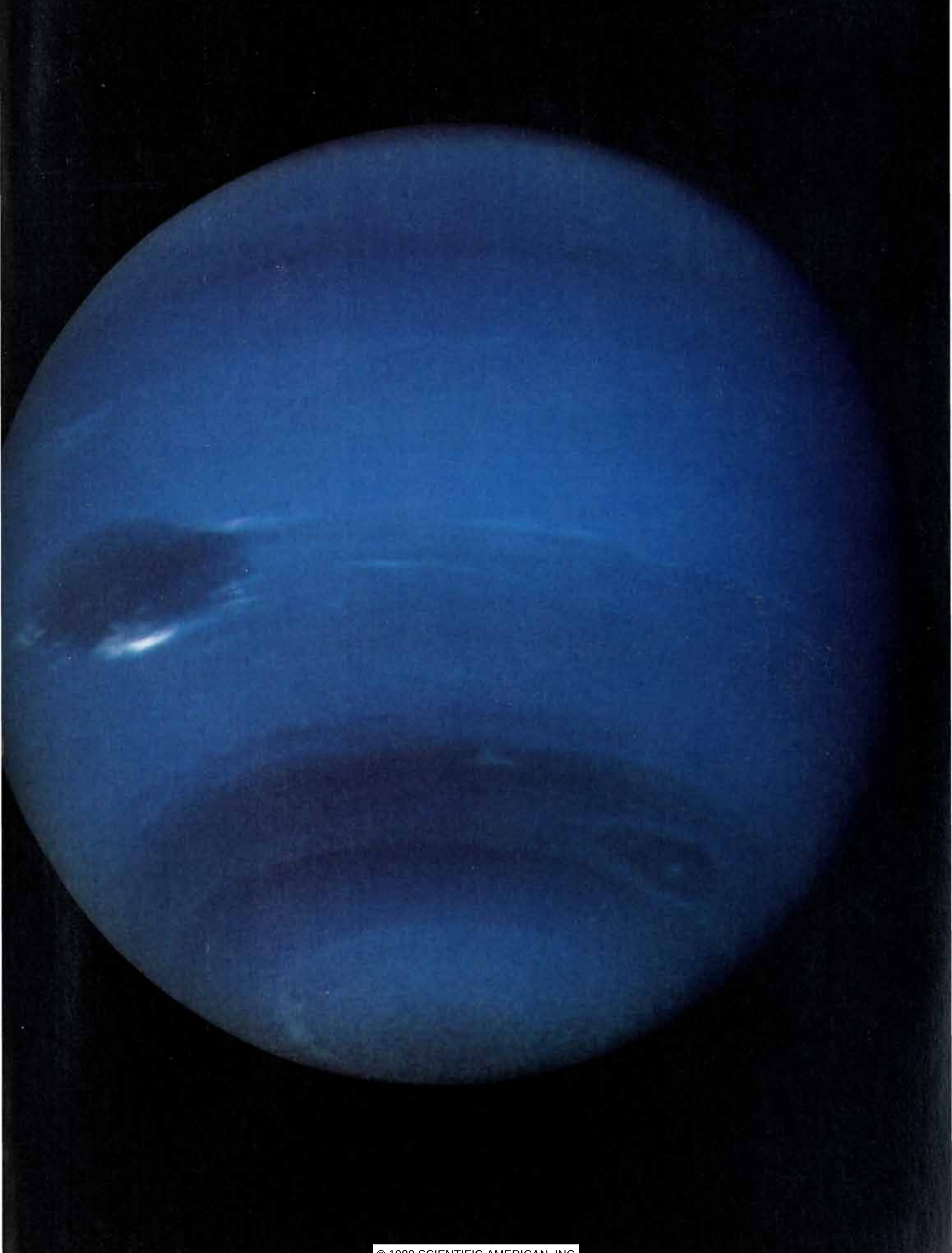
Voyager 2's cameras unveil a stormy world and a frozen moon molded by volcanism

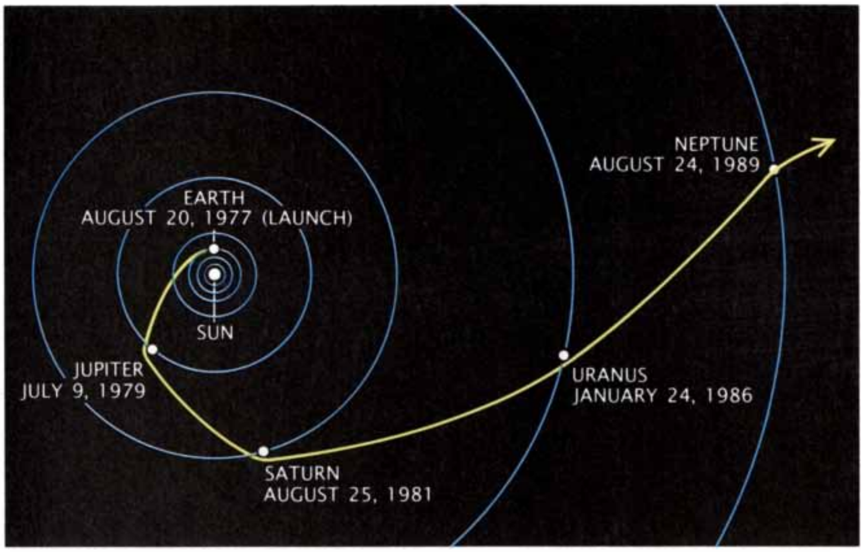
by June Kinoshita

On the night of August 24, a small, angular contraption hurtled over the cloud tops of Neptune. It swooped barely 3,000 miles above the great blue planet's north pole, plunged down the night side, scooted past the large moon Triton at a distance of 24,000 miles and vanished into the void. During that brief encounter the visitor meticulously snapped thousands of images and radioed them to the earth. Scientists waiting at the Jet Propulsion Laboratory in Pasadena, Calif., cheered and uncorked the champagne as the pictures—humankind's first close-up look at the eighth planet—came into focus on their screens.

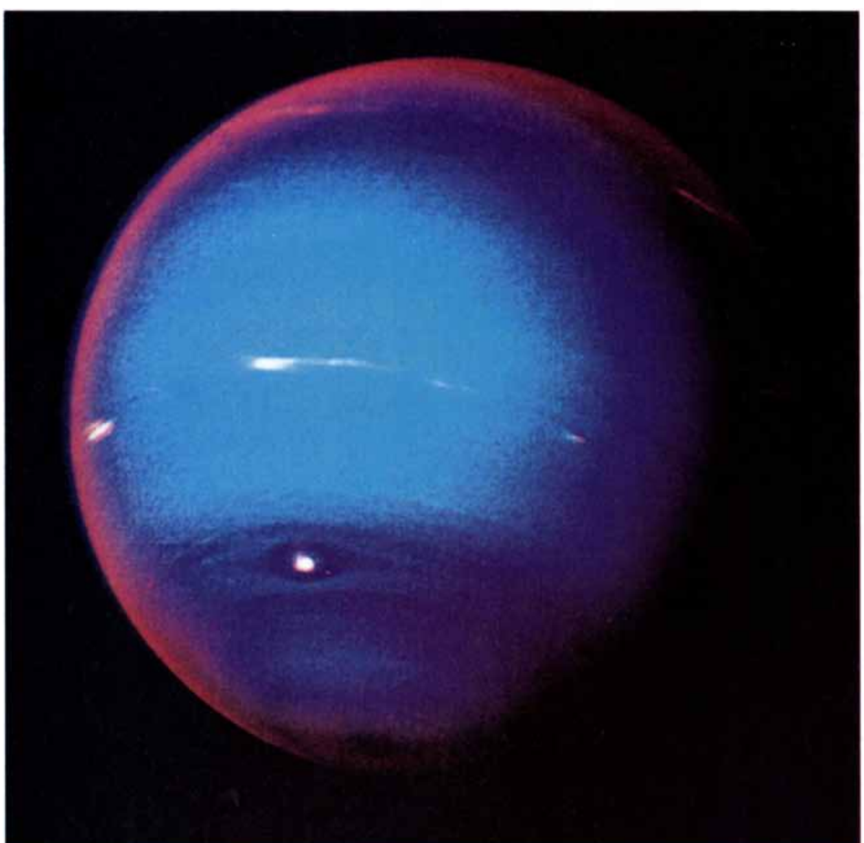
Voyager 2 had taken 12 years to reach Neptune, the fourth and final destination of a planetary pilgrimage that included Jupiter and Saturn (both were also visited by the probe's twin, *Voyager 1*), as well as Uranus. Of all the planets on the itinerary, Neptune was the least known. Overhauling the on-board computer programs and gingerly firing the thruster rockets, the Voyager team steered the aging ship to a flawless encounter. From signals that reached the earth with a strength of less than a ten-quadrillionth of a watt, the team gleaned images of breathtaking clarity. On these pages *Scientific American* presents the last fruits of the epic journey.

BLUE, STORM-STREAKED ORB of Neptune was captured in this image made from a distance of 6.6 million kilometers **(right).** ***A violent geologic past scarred the face of Triton, Neptune's largest moon*** **(below).**





VOYAGER'S GRAND TOUR of the outer planets took advantage of a once-in-a-176-year planetary arrangement that enabled the spacecraft to be catapulted from one planet to the next by gravity. Originally designed to probe only Jupiter and Saturn, **Voyager 2** ***was remotely reprogrammed to operate at Uranus and then Neptune, currently the most distant planet in the solar system. (Pluto's eccentric orbit has for the moment taken it inside Neptune's orbit.) The spacecraft dove over the north pole of Neptune, sped past Triton and headed south out of the solar system.***



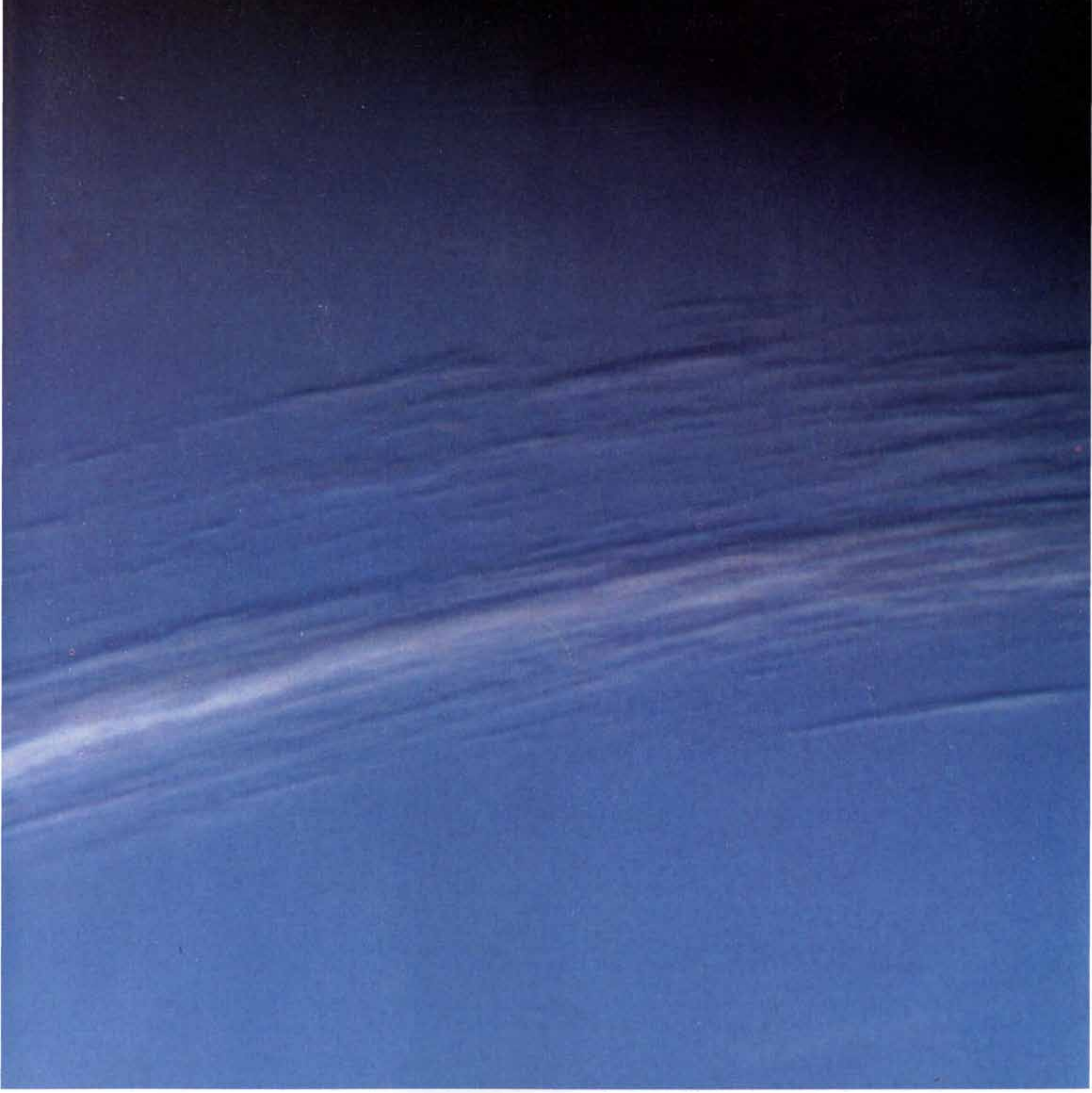
METHANE GAS in Neptune's atmosphere is mapped in this false-color image made with a filter that admits light at a wavelength absorbed by methane gas. Haze high above the methane layer reflects sunlight at the edge of the disk, resulting in a bright red rim. At the center of the disk, sunlight penetrates the haze and is absorbed by the methane, giving rise to a blue color. Highly reflective cirrus clouds appear as bright patches of white.

Orbiting at a distance of 4.5 billion kilometers from the sun, where the light is a thousandth as strong as it is at the earth, Neptune barely shows up as a pale-green speck in the most powerful earthbound telescopes. Indeed, the planet was found only 143 years ago, after astronomers speculated that the gravitational pull of an eighth planet accounted for anomalies in the orbit of Uranus. Since then, observers have estimated Neptune's mass, size and composition, all of which suggested that it would be much like its bland "twin," Uranus. To their surprise, *Voyager 2* revealed a turbulent world, with giant storm systems rivaling those of Jupiter and fleeting clouds unlike any seen before on a gaseous planet.

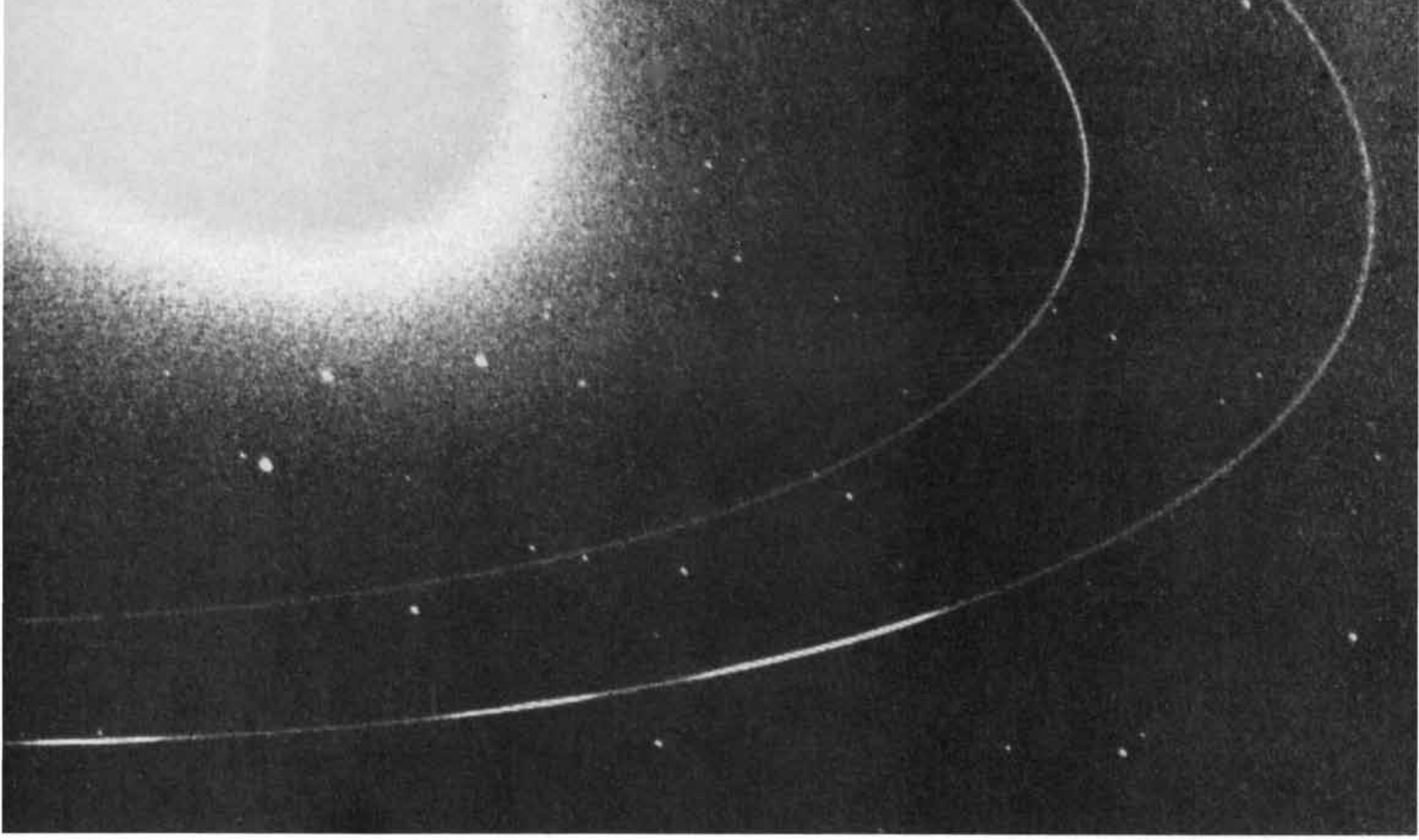
Like Uranus, Neptune is a great ball of water and molten rock cloaked in an atmosphere of hydrogen and helium mixed with methane. The methane absorbs red light and is responsible for the planet's aqua hue. Unlike Uranus, Neptune's atmosphere bears distinctive striations and gigantic dark storm spots, including one hurricane as wide across as the earth, which scientists quickly christened the Great Dark Spot. First detected by *Voyager 2* this past winter, the massive storm is located at about 22 degrees south latitude and appears to churn in a counterclockwise direction. Later, the cameras picked up a smaller dark spot farther south. *Voyager* images also showed a small bright cloud, named Scooter, dashing along at an intermediate latitude between the two spots.

Radio signals, which issue from the planet like a beam from a lighthouse, disclosed that Neptune completes one rotation in 16 hours and three minutes—about one hour faster than had been predicted. Images made over several complete rotations revealed that the small dark spot travels at about the same rotational period and so remains in roughly the same position with respect to the planet. The Great Dark Spot takes longer, about 18 hours, to complete a rotation and so must be sweeping westward against the planet's rotation at a speed of 300 meters per second, or 700 miles per hour, driven by the fastest retrograde winds *Voyager 2* has ever clocked.

Two hours before *Voyager 2*'s closest approach to the planet, the spacecraft's cameras happened on a startling and beautiful sight: catching the oblique rays of the sun, parallel banks of silvery cirrus clouds were casting shadows on the blue cloud deck below. From the po-



SILVERY CIRRUS CLOUDS in Neptune's northern hemisphere cast shadows onto the blue cloud deck 50 kilometers below. The clouds extend over thousands of kilometers. Two images taken 17.6 hours apart from a distance of 12 million kilometers document the atmosphere's dynamism (inset). *The planet completed a little more than one full revolution in the interim; the small dark spot rotated at the same speed as the planet, but the Great Dark Spot swept westward at a speed of 1,000 kilometers per hour.*



"LOST ARCS" turned out to be bright clumps in Neptune's outer ring, seen here from a distance of 1.1 million kilometers as Voyager 2 *left the planet.*

sition of the shadows and the angle of the sun, members of the imaging team estimated that the clouds hovered some 50 kilometers above the underlying layer. Scientists were ecstatic: such three-dimensional structures had never before been observed in the atmosphere of any other giant planet.

The delicate wisps testified, paradoxically, to a dynamic atmosphere. According to Robert West of the photopolarimetry team, the stratification indicated that Neptune is more dynamic than even Jupiter, which, apart from its roiling red spot, has a flat if colorful cloud layer. Scientists cannot yet explain how the high clouds form or why they extend along only a few lines of latitude.

High cirrus clouds also draped the south rim of the Great Dark Spot and formed a bright dimple above the center of the small dark spot. The cloud formations remained in more or less the same location, even though they were surrounded by violent winds. Bradford A. Smith, head of the imaging team, speculated that updrafts loft gaseous methane high into the atmosphere, where it freezes into icy cloud particles; downdrafts then drag the clouds down to warmer regions where they dissipate. A similar process creates cloud formations above terrestrial mountains.

Neptune's magnetic field held surprises: its dipole axis is tilted some 50 degrees from the rotation axis and is also displaced from the planet's center by 10,000 kilometers. The finding helped to clarify a problem that had puzzled planetary scientists ever since *Voyager 2* disclosed that Uranus's magnetic-field axis was similarly tilted. Because the magnetic fields of other planets tend to coincide with their rotation axes, scientists had speculated that the tilt might be related to the unique orientation of Uranus's rotation axis, which lies in the plane of its orbit. Another possibility was that, by sheer chance, *Voyager 2* had caught the planet in the middle of a reversal in the direction of its magnetic field.

But with the new finding about Neptune's magnetic field, neither explanation is plausible anymore. Neptune's rotation axis has the usual more nearly perpendicular orientation, and the chances of catching two planets in the midst of magnetic reversals are slim. On the other hand, according to Norman F. Ness, head of the magnetic-field experiment, skewed magnetic fields are common to a class of stars called oblique rotators. He suggested the planets' fields could arise by the same mechanism that has been proposed for the stars: the convection of electrically conducting material within a thin, spherical shell near the surface. (On the earth the convection occurs within a molten metallic core.)

The tilted field threw a wrench into some of *Voyager 2*'s other experiments. Edward C. Stone, chief project scientist of the Voyager mission, said navigators had hoped that by aiming for Neptune's geographic north pole, they would send the spacecraft across the converging magnetic-field lines of the auroral zone usually associated with the magnetic pole of a planet. Instead it entered the planet's magnetosphere—the ion-rich envelope created by the planet's magnetic field—along the converging lines, following them down toward the magnetic pole. This was dumb luck: according to Stone, no probe has ever flown that route before on any other planet, including the earth.

Voyager 2 did sight auroras in Neptune's atmosphere, but they were spread over a wide region rather than forming well-defined ovals around the magnetic poles, as on the earth. Auroras were also seen on Triton. Andrew Cheng of the Low-Energy Charged-Particle Team reported that charged particles in Neptune's radiation belts appear to plunge into Triton's atmosphere with sufficient energy to generate the ultraviolet auroras that were observed there.

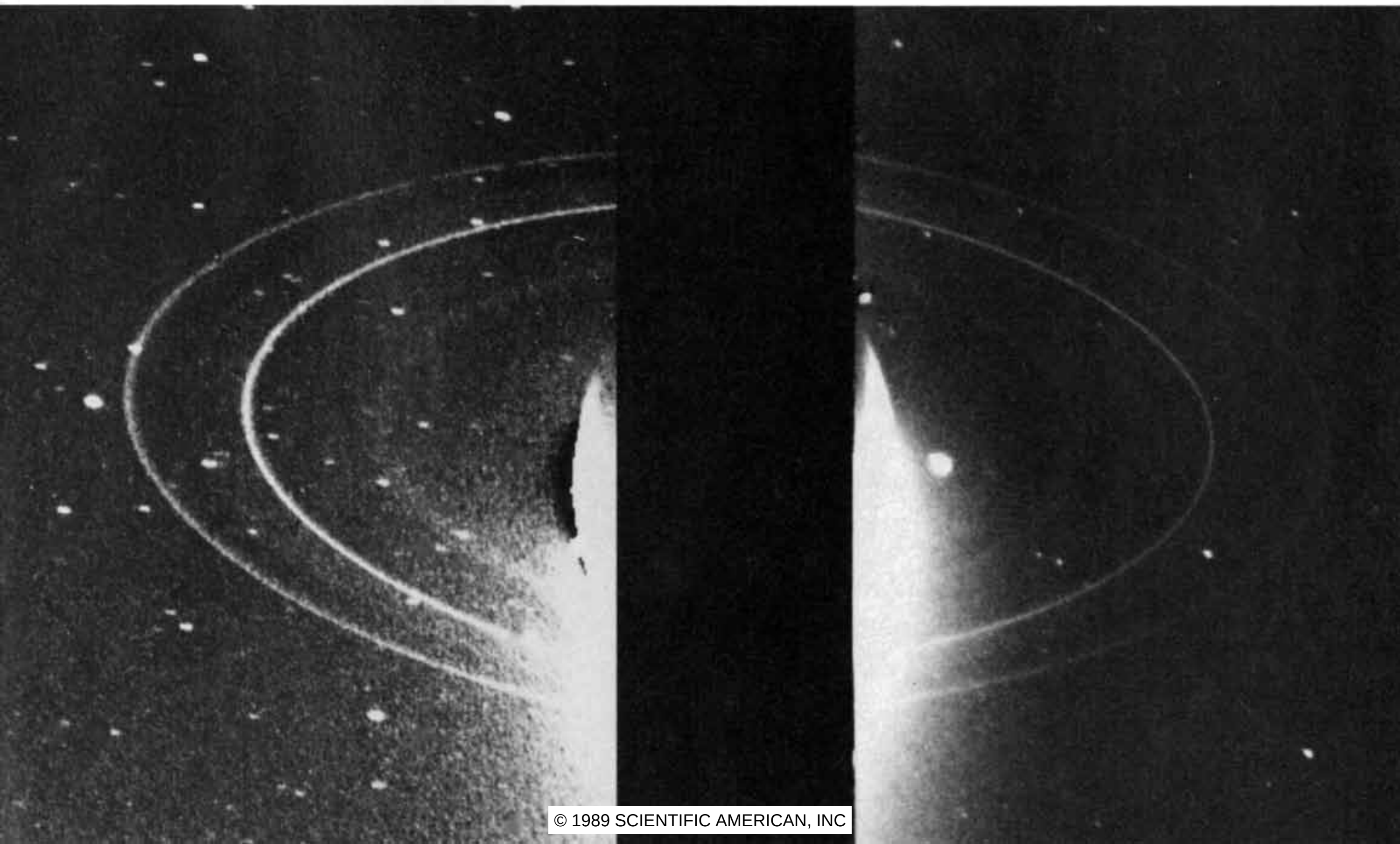
In a mission filled with unexpected twists, one of the most dramatic was the resolution of the quest for the "lost arcs." Back in 1984 telescopes on the earth detected what appeared to be incomplete rings, or ring arcs, around Neptune. If partial rings indeed existed, they would be the first to be seen around any planet. In early August *Voyager 2* seemed to confirm the presence of arcs, but as the spacecraft closed in on its target it began to detect wispy segments between the arcs, and by August 24 the imaging team announced that the arcs were part of a complete outer ring. A remarkable time-lapse image revealed clumps of fine dust no more than 10 or 20 kilometers across embedded in one arc. Scientists have so far failed to come up with a plausible explanation for how the arcs formed. Other images revealed that Neptune has four rings altogether. The dust making up the rings is thought to be the debris thrown off over the eons when micrometeorites smashed into the moons.

Observations of starlight being occulted by the outer ring indicated that it has a dense core about 17 kilometers across, surrounded by a diffuse halo of dust about 50 kilometers across. As *Voyager 2* crossed the plane of the ring about one hour before its closest approach to the planet, the plasma-wave detector transmitted a barrage of radio pulses generated by dust particles as they struck the spacecraft and vaporized into microscopic puffs of plasma. The hailstorm of pulses played back on audiotape the following morning by Donald J. Gurnett, head of the plasma-wave team, indicated as many as 300 strikes per second, or one particle per 300 cubic meters—comparable to the dust in Saturn's ring plane.

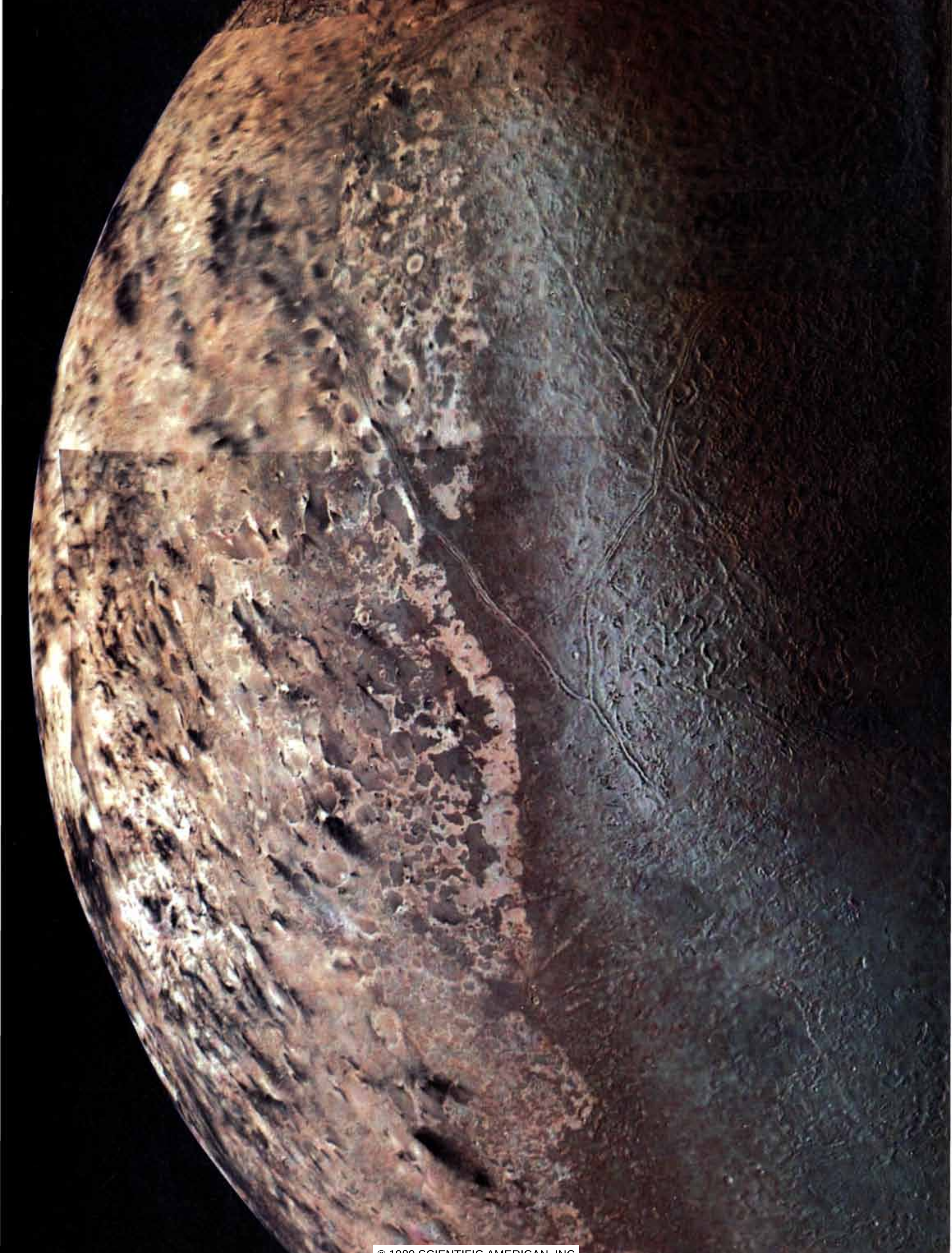
Two of Neptune's moons, Triton and Nereid, were known from terrestrial observations. *Voyager 2* found six new satellites, temporarily catalogued as 1989 N1 through 1989 N6. These dark, misshapen chunks range from 50 to 200 kilometers in diameter—too small to pull themselves into a sphere through gravity. *Voyager*'s grainy images of the meteor-battered bodies indicate that the small moons have remained essentially unmelted since the time they were formed.

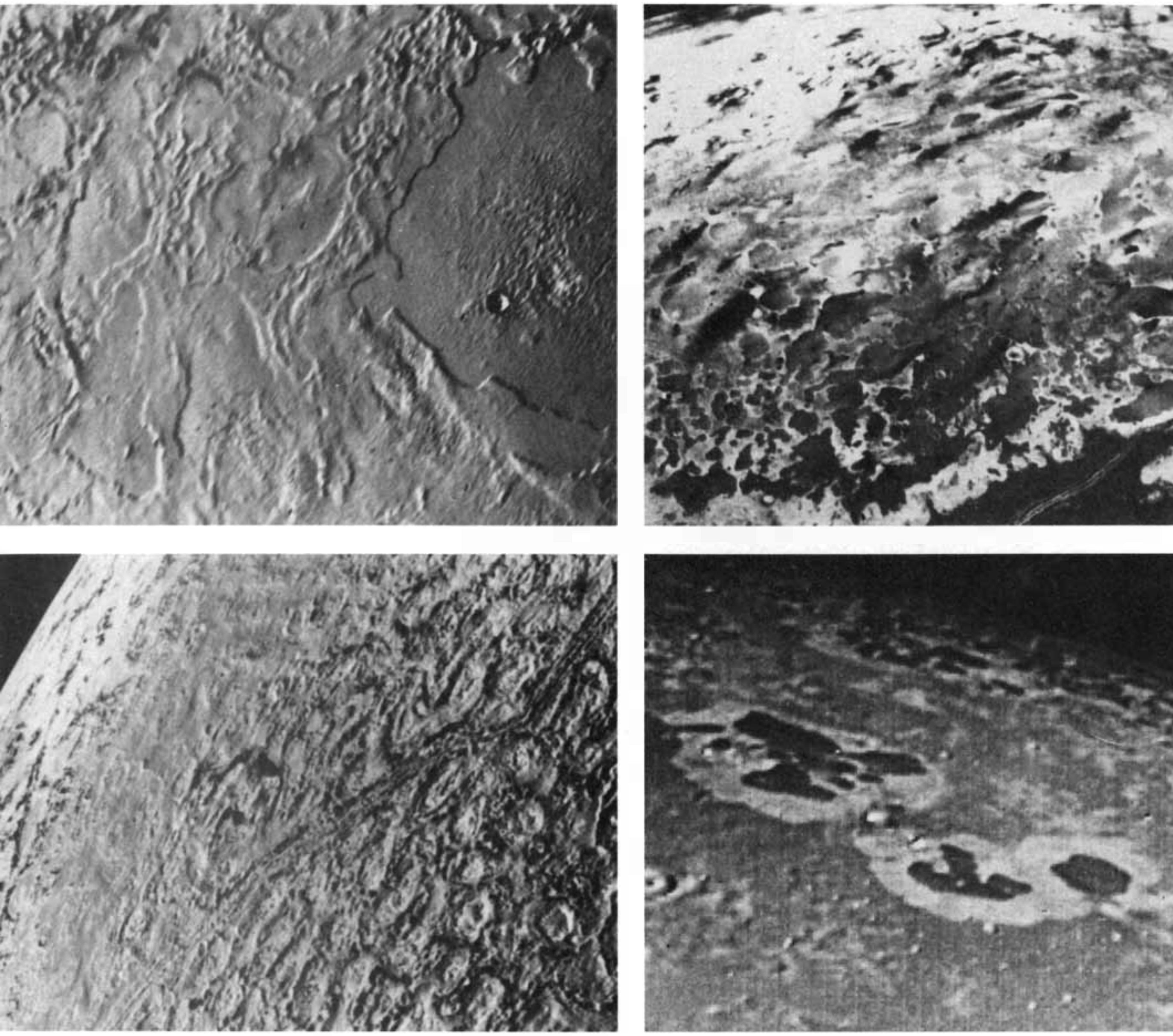
All of the newly found moons orbit near the equatorial plane of Neptune and in the same direction as the planet rotates. The orbital planes of Triton and Nereid, in contrast, are tilted by 20 and 30 degrees, respectively, and Triton orbits in the retrograde direction—the only large moon in the solar system to do so. The sharply tilted orbits suggest that these moons did not condense from the same matter as the planet but instead are foreign bodies that tumbled into Neptune's gravitational embrace. Dale P. Cruikshank of the Infrared Radiometry and Spectrometry Team observed that Nereid closely resembles a distant asteroid called Chiron. "Nereid could have been a cousin to Chiron," he said. "These could be the planetesimals from which the planets evolved."

If scientists knew little about Neptune before the *Voyager 2* flyby, they knew even less about its largest moon, Triton. As one of the Triton investigators put it: "We knew it was there, and we knew what it was called." In the days leading up to the encounter, *Voyager 2* transmitted teasing snapshots of Triton: a bruised, pinkish ball hinting at a spectacular geologic history. On the night of August 24, scientists stocked up on tortilla chips and coffee to sustain them until dawn, when the images from



COMPLETE RINGS can be seen in a pair of magnificent backlit images. There are two bright rings, a fainter inner ring and a diffuse sheet that may descend to Neptune's cloud tops. The planet itself was left out, accounting for the black strip in the middle.





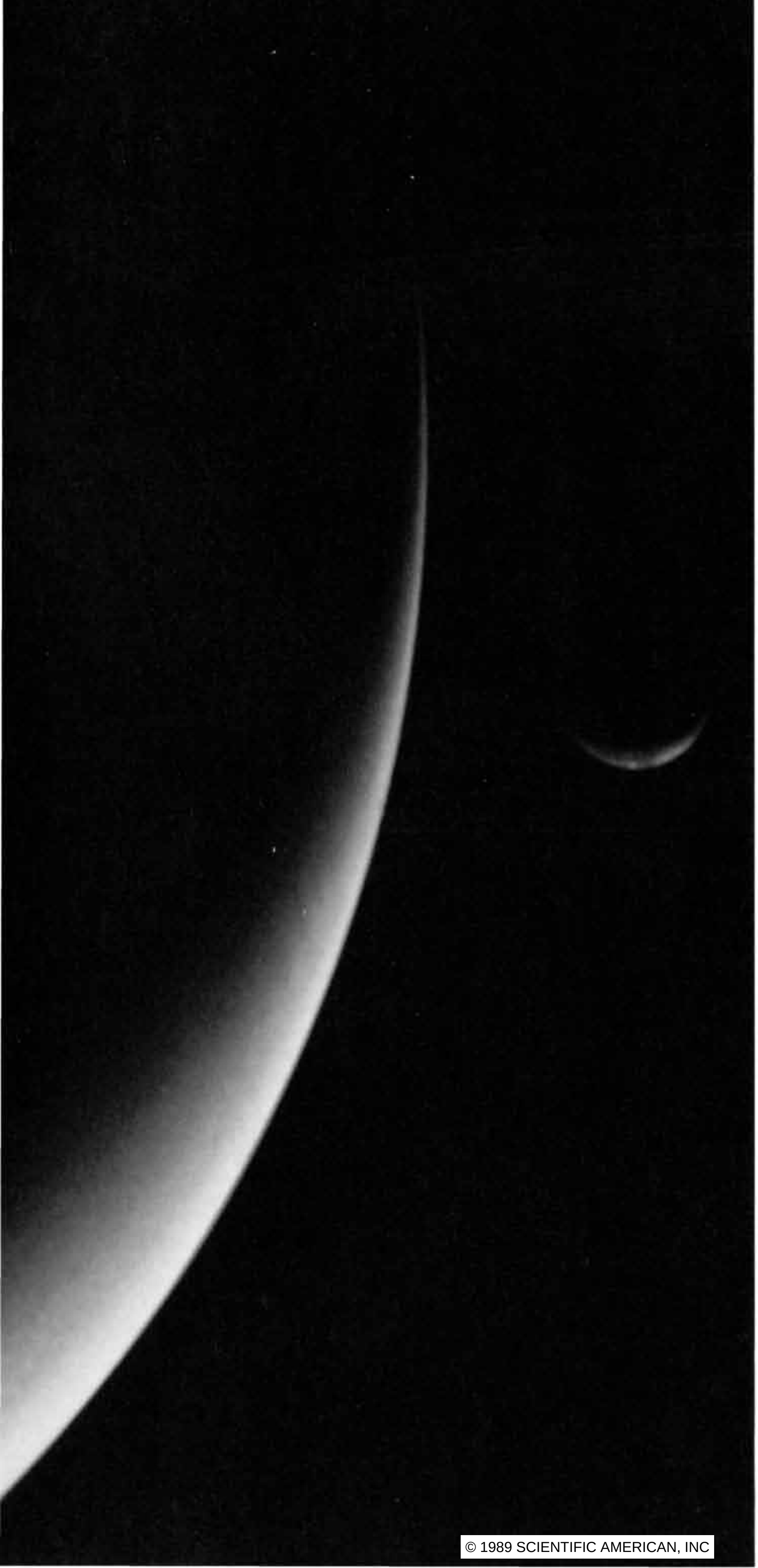
"A WORLD UNLIKE ANY OTHER" is how scientists described the tortured moon Triton, seen in a composite of a dozen high-resolution images (opposite page). *The large south polar cap on the left side of the image may consist of a slowly evaporating crust of nitrogen ice deposited during the previous winter, some 80 years ago. Gigantic fissures scar the moon's face. Frozen lakes* (top left) *may have been melted by volcanism. Plumes of dark material over the polar frost may be evidence of more recent volcanism* (top right). *Cantaloupe-like terrain* (bottom left) *may have been created by local melting and deformation of the surface. Dark patches surrounded by a bright rim* (bottom right) *leave scientists still stumped.*

the spacecraft's closest approach to Triton would reach the earth. In the pressroom, reporters slumped over typewriters or paced expectantly in front of the monitors.

At 3:40 A.M., the first of the black-and-white images flashed on the screens. People leapt to their feet and rushed up to the monitors, exclaiming and pointing as the remarkably sharp pictures disclosed crenulated landscapes, vast canyons, craters and peaks. Each new image was more bizarre than the last. "It looks like tripe," someone said. Others suggested "cantaloupe" and "cellulite." It was, as a tired, happy Ed Stone remarked later that morning, "a night to remember."

Triton proved to be 2,720 kilometers in diameter—somewhat smaller than the earth's moon. Ices of methane and nitrogen encrust its polar cap, which reflects away so much sunlight that the temperature is a mere 37 kelvins, making Triton "the coldest object we have seen in the solar system," said Roger Yelle of the ultraviolet spectrometry team. Even so, the moon's tilted rotational axis and tilted orbital plane give rise to extremes of season. The south polar cap, now at the height of its 41-year-long summer, has evaporated away at many spots along its border.

Triton's atmosphere is exceedingly thin, about 100,000 times thinner than the earth's, and consists mainly of nitrogen. A magnified view of the moon's limb (its silhouetted edge) revealed that the atmosphere, thin as it is, is still sufficient to support a haze of small particles from about five to 10 kilometers above the surface. The atmosphere's temperature rises to about 100 kelvins at an altitude of 600 kilometers. Scientists say the phenom-



VOYAGER'S LAST LOOK at Neptune caught the crescent of Triton hovering near the looming belly of the planet. The image was taken three days after the closest approach, as the spacecraft was plunging southward out of the solar system.

enon is unlike a temperature inversion in the earth's atmosphere. The heating in Triton's atmosphere is occurring at a higher altitude, and so far no one understands what could be causing it.

Much curiosity centered on the dark smudges resembling wind streaks on Triton's south polar cap. Laurence A. Soderblom of the imaging team raised one of the biggest stirs of the encounter by suggesting that the streaks were caused by volcanic eruptions or geysers. Temperatures and pressures near Triton's surface would allow some internal heat source to raise subsurface liquid nitrogen to pressures at which it would explode, he said. As the nitrogen was ejected into the atmosphere, it might have swept along darker carbon compounds from the crust. The dust could have wafted downwind and blown onto the frost. Soderblom suggested that such eruptions might be occurring all over the moon and that it is the icy ground cover that makes the plumes visible only on the south polar cap. Similar plumes of sulfur dioxide were seen on the Jovian moon Io, Soderblom noted. "This is a crazy idea, and it's probably wrong," he said, "but it's the best we have in the meantime."

Just to the north of the polar cap lies a vast tract of evenly sized circular depressions and ridges that resembles nothing so much as the rind of a cantaloupe. "The area has been faulted and deformed an indescribable number of times," Soderblom said. The surface bears noticeably fewer craters than an adjacent region, which indicates that it is probably the youngest terrain on the moon. Fissures slash the surface like superhighways, meeting in gigantic X's and Y's. Viscous material, perhaps a slush of water laced with ammonia, appears to have forced its way up into some of these fissures, forming central ridges and in some cases overflowing onto the surrounding plain.

Lying within this terrain are frozen lakes edged with a series of terraces, as though the level of the lakes had changed as a result of repeated freezing and melting driven by volcanic heat. Such terraced formations are common in Hawaiian volcanoes, Soderblom said. The lakes on Triton must have once been filled with a low-viscosity liquid, because their frozen surface is level. Yet in frozen form the substance must be extremely rigid in order to sustain kilometer-high terraces. Methane, nitrogen and carbon monoxide are implausible candidates, said Steven K. Croft, a geologist at the University of Arizona, because in their frozen form they would flow like glaciers. Water ice, however, is as rigid as rock at the temperatures found on Triton and is "almost certainly" the material out of which the lakes formed, he said.

These signs of past volcanism prove that Triton was once a hotter place, presumably because it had an unusual origin. Triton may once have been an independent planet rather like Pluto, which it resembles in size and possibly composition—it contains more rock than do the other icy satellites. Later, Triton was captured by Neptune; as it gradually settled into its current circular orbit, tidal friction would have melted the moon and driven the volcanism until perhaps from one to two billion years ago. "It's likely that we're seeing the frozen imprint of that earlier era," Stone said.

Voyager's reconnaissance of Triton ended an extraordinary epoch of planetary exploration. In the 12 years since they were launched, the *Voyager* spacecrafts have contributed more to the understanding of the planets than three millennia of earthbound observations (and at a cost of only $2.40 per U.S. citizen for the entire show). Science now has "nothing less than the encyclopedia of the planets," Stone exulted.

Voyager 2 is now heading southward out of the plane of the planets at an angle of about 50 degrees. *Voyager 1* veered northward after the Saturn flyby. The spacecrafts' plutonium power sources are expected to sputter out around the year 2015. By that time scientists hope the *Voyagers* will have encountered the heliopause—the true edge of the solar system, where the solar wind collides with the interstellar medium. They will then drift sightless and voiceless through the eons, a testament to the questing spirit of the humans who launched them.

JUNE KINOSHITA is International Coordinating Editor and a member of the Editorial Board of *Scientific American.* She thanks Edward C. Stone, professor of physics at the California Institute of Technology and chief project scientist for the Voyager mission, for his assistance in preparing this article.

Shuttle Glow

Objects in low earth orbit travel through a tenuous and yet highly reactive atmosphere. Some materials barely survive the trip; others glow along the way

by Donald E. Hunton

Where does the earth's atmosphere stop and space begin? For most of us the answer would certainly be below the domain of orbiting satellites and other spacecraft. After all, why would the National Aeronautics and Space Administration call their newest vehicle the space shuttle if it did not go into space?

In fact, the question of where the atmosphere ends is about as meaningful as "How high is the sky?" The pressure of the earth's atmosphere decreases smoothly with increasing altitude, starting with the familiar "one atmosphere" at sea level and ending with the near-perfect vacuum of interplanetary space, beyond the orbit of the moon. There is no precise point where the atmosphere ends and space begins.

The space shuttle and many other satellites orbit at altitudes of 250 to 300 kilometers, only about 25 times as high as commercial jet aircraft fly. The atmosphere at this altitude is far from a perfect vacuum: it contains about a billion (10^9) particles per cubic centimeter, in comparison with about 3×10^{19} at sea level. It is still very much a part of the layer of gases that surrounds our planet.

The space shuttle flies through these thin gases at a velocity of eight kilometers per second, which causes the air to rush past the spacecraft. From the shuttle a strong wind seems to be blowing, analogous to the wind a person might feel when putting a hand out the window of a speeding car. (The force, however, is minuscule: only one millionth the force of gravity.) The impact of this wind on the shuttle's surface gives rise to two remarkable phenomena: the spacecraft glows in orbit, and its body is slowly eroded away.

During the third shuttle flight, in March, 1982, Peter M. Banks and his co-workers at Stanford University and Utah State University asked shuttle astronauts Jack R. Lousma and C. Gordon Fullerton to take photographs of an experiment they had placed in the payload bay. In addition to recording the results of the experiment, the pictures showed an unexpected faint orange glow above the shuttle's tail and engine pods where they faced directly into the orbital wind. The light came not from the shuttle itself but rather from a 20-centimeter-thick region directly above its surface. The glow was sometimes visible to the naked eye and could be easily recorded by making long exposures with an ordinary camera and film.

Meanwhile engineers examining the shuttle after its early flights observed significant changes in the condition of many of the materials that made up the craft and its payloads. Kapton, a plastic film used extensively for thermal insulation, is generally translucent, glossy and deep amber in color. After spaceflight, some samples were dull, light yellow and opaque. The carbon coating on two metal spheres that served as sensors for electric-field measurements had been eroded away completely. Silver films were badly oxidized. Thermal-control paints had lost their gloss.

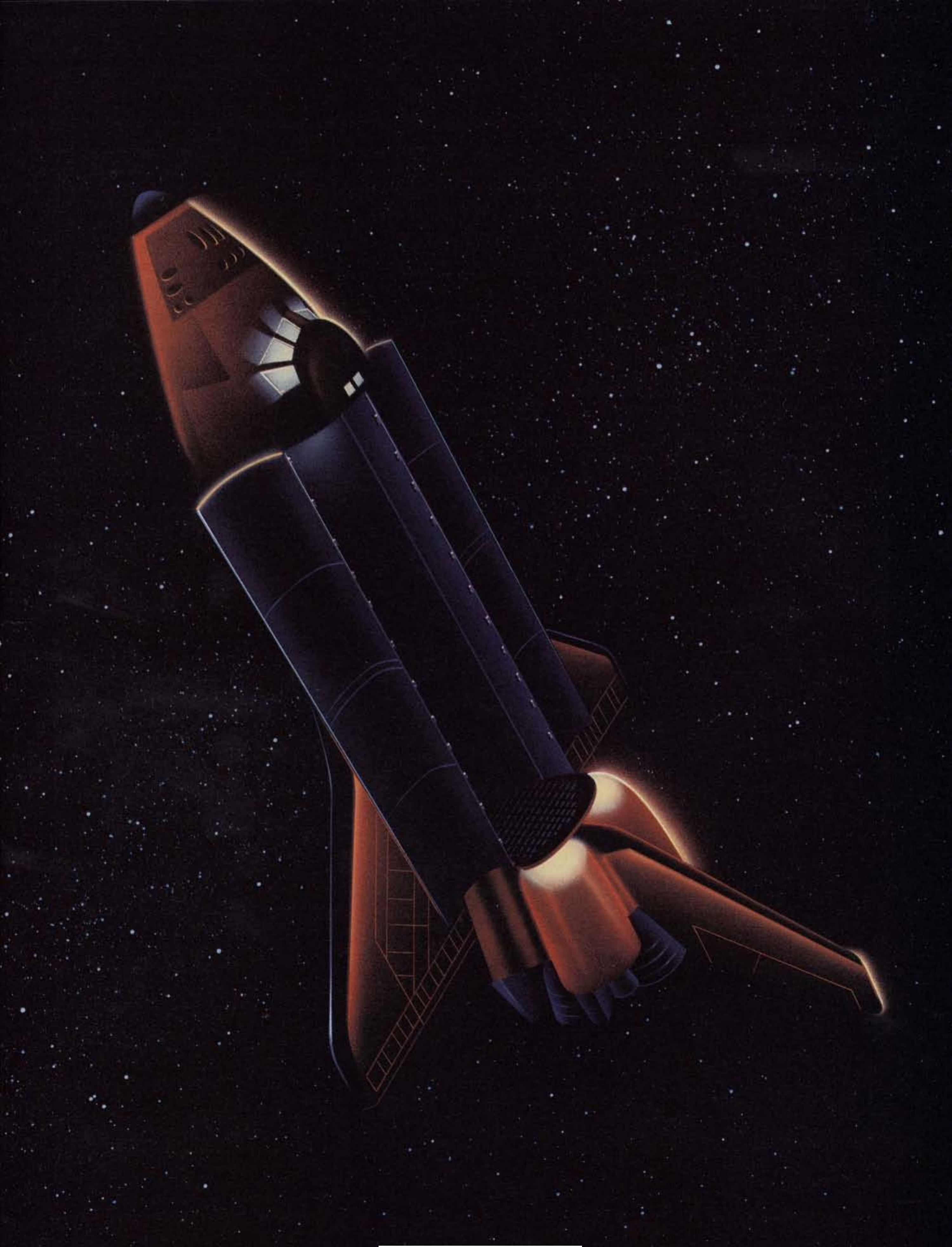
The twin phenomena of material erosion and shuttle glow have profound implications for the design of future spacecraft, especially ones that will spend decades in low earth orbit, such as the space station. The phenomena are now the subject of intense research. Neither glow nor erosion is now fully understood, but together they testify to the highly reactive nature of the uppermost atmosphere. Additional experiments to clarify their causes and effects will fly on board the shuttle in the next few years.

Several important differences between the earth's atmosphere at ground level and at shuttle altitudes help to create an environment where glow and surface erosion can occur. The pressure at 300 kilometers is approximately one ten-billionth of the pressure at ground level but is still between 10 million and 100 million times the pressure in interplanetary space. The chemical composition of the uppermost atmosphere also differs substantially from what is found near ground level. Above about 100 kilometers the familiar mixture of about 20 percent oxygen, 80 percent nitrogen and various trace gases disappears and is replaced by a much more reactive composition.

One important contributor to this change in composition at high altitudes is visible and ultraviolet light from the sun. Much of this solar radiation is absorbed before it reaches the ground; oxygen molecules in the upper atmosphere can intercept enough ultraviolet light to break apart into two oxygen atoms. As a result, roughly 80 percent of the atmosphere at shut-

DONALD E. HUNTON is a physicist in the Air Force's Geophysics Laboratory at Hanscom Air Force Base in Massachusetts. He studies shuttle glow, surface erosion and other phenomena that occur in high-temperature, low-density gases and plasmas. He studied chemistry as an undergraduate at Dartmouth College and received his doctorate in chemical physics from the University of Colorado at Boulder in 1983. Hunton is collaborating with investigators from the National Aeronautics and Space Administration and other national space agencies on further investigations of shuttle glow; one of his experiments is scheduled to fly in late 1990.

FAINT GLOW surrounds the space shuttle as it plows through the tenuous atmosphere of atomic oxygen and molecular nitrogen 250 kilometers above the earth's surface. This hypothetical view from outside the shuttle was drawn on the basis of photographs made from the spacecraft's aft flight deck and analysis of mechanisms responsible for the glow.



tle altitudes consists of atomic oxygen. Absorption of solar radiation also increases the temperature of the uppermost atmosphere to about 800 degrees Celsius.

The hot atomic oxygen is highly reactive and is thought to be responsible for both erosion and glow. At sea level, molecular oxygen is responsible for the corrosion and eventual destruction of metals and other materials; at orbital altitudes, atomic oxygen participates in the same kinds of chemistry, but it acts much more fiercely.

When the erosion of shuttle surfaces was first discovered, a group of NASA engineers and scientists led by Lubert J. Leger proposed that the phenomenon was caused by reactive atomic oxygen. Erosion, the researchers hypothesized, was analogous to sandblasting on a molecular scale: fast-moving atmospheric gases were etching away plastics and other materials by means of chemical reactions. They quickly planned a series of careful experiments for later shuttle flights to test their hypothesis.

These experiments were flown on board shuttle missions in November, 1982, and September, 1983. They have yielded most of the quantitative data available on the effects of the orbital environment on surfaces and materials. Samples of a wide variety of materials found in spacecraft were mounted in the shuttle's payload bay; in flight the opened bay was oriented so that the flow of atomic oxygen struck the samples directly for 40 hours.

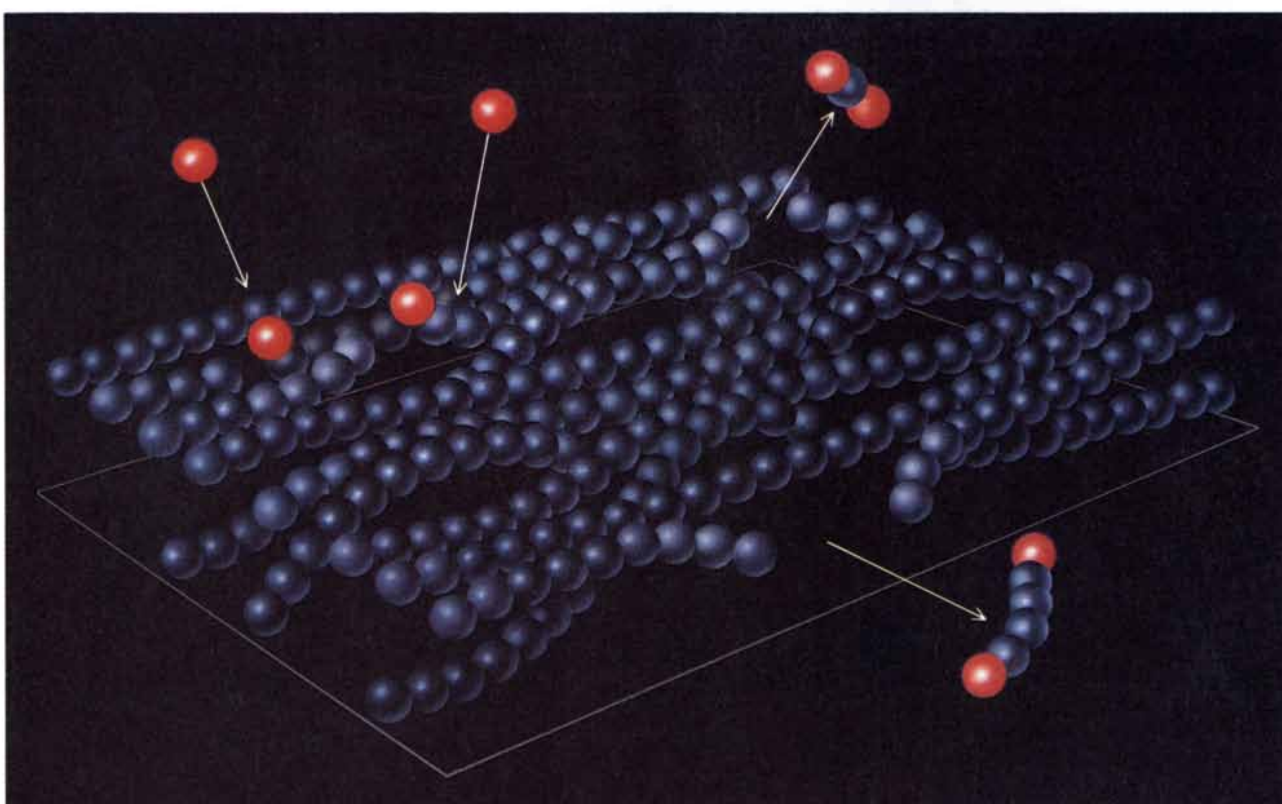
After the flight the samples were returned to the laboratory, where they were weighed to determine how much mass they had lost and then photographed at high magnification under a scanning electron microscope. The profiles of the samples were measured with high accuracy to determine how thick a layer had been etched away, and each sample's composition was determined in search of possible chemical changes.

The findings of these experiments were just as dramatic as the early observations. The mass and thickness of many samples decreased as they were eaten away by the atomic oxygen. Exposure had left the surfaces uneven, deeply pitted and rough.

In general, plastic and other organic materials (those containing only carbon, hydrogen, oxygen and nitrogen atoms) were eroded much more dramatically by atomic oxygen than metals were. The reaction rate, however, was essentially independent of the detailed chemical structure of the materials; almost all polymers were eroded at about the same rate. Fluorinated polymers such as Teflon were among the few nonmetals that did not react strongly with the atomic oxygen.

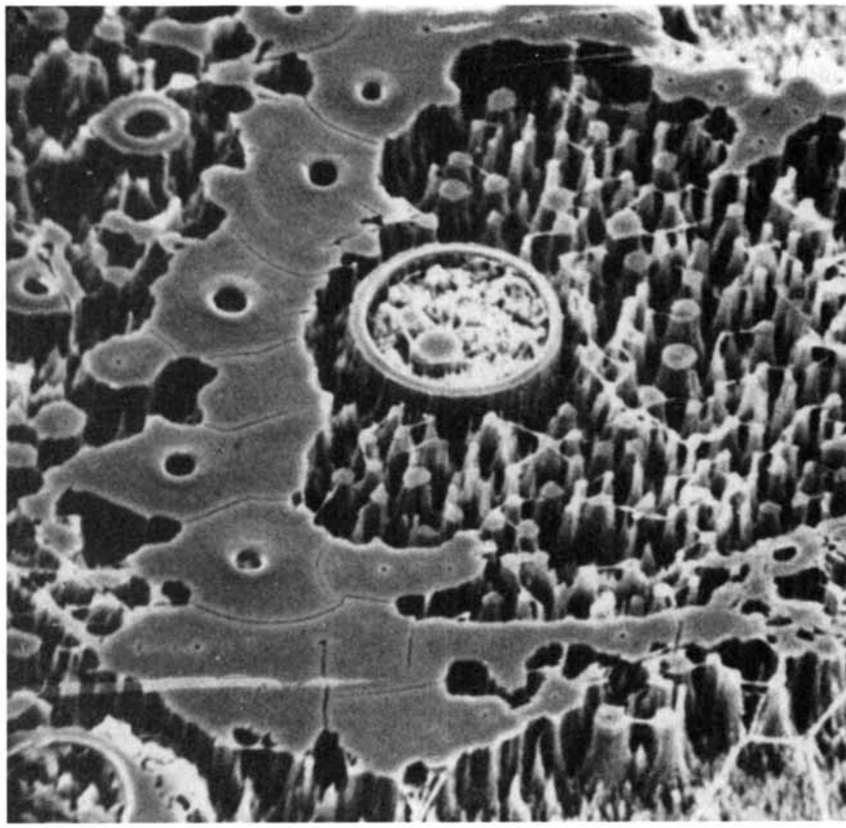
The most reactive organic materials lost about 12 microns (millionths of a meter) of thickness during the 40-hour exposure in orbit—about one fifth the thickness of the paper this article is printed on. Even though 12 microns in 40 hours seems a relatively low rate of erosion, it could have catastrophic effects on very thin films. A typical antireflection coating on a lens, for example, is two or three microns thick at most. A lens directly exposed to the orbital flow of oxygen might lose its coating in a matter of hours. During the course of a year in orbit at the altitude of the experiments, a layer of material 2.5 millimeters thick could be eroded away completely.

The details of the erosion mecha-



ATOMIC OXYGEN ATTACKS spacecraft surfaces in orbit. Polymers composed largely of hydrocarbon chains, such as polyethylene and Kapton, are particularly vulnerable. Atomic oxygen (*red*) from the upper atmosphere strikes the surface at eight kilometers per second, breaking a chain by bonding to a carbon atom (*black*) at (*left*). The resulting shortened chain (*center*) is later struck by another oxygen atom, yielding a molecular fragment (*right*) that leaves the surface.

nism are not clear, but as Leger and his colleagues proposed, the most likely process involves chemical attack by oxygen on the surface material. Atomic oxygen may break the long chains of the polymer molecules into shorter pieces that can escape from the surface. The reaction could also break down the organic material into simple molecules such as water or carbon dioxide. Workers at a number of laboratories are bombarding samples with beams of atomic oxygen under conditions intended to match those in orbit. Their goal is to determine the mechanisms responsible for erosion and to test the erosion resistance of new materials and protective coatings without sending them into orbit. Recent experiments have demonstrated effects very similar to those observed after shuttle flights.



POLYMER SURFACE was eroded by atomic oxygen during spaceflight. This sample of Kapton, which is magnified approximately 10,000 times in a scanning electron micrograph, had been sprayed with a protective silicone coating, but the coating was too thin to wet the surface uniformly. The Kapton film was preserved only in certain regions; unprotected regions were left roughened, and their thickness was reduced.

Experiments to help determine the cause of shuttle glow were formulated with similar alacrity. Researchers studying the phenomenon quickly agreed that the light must come from excited molecules near the surfaces struck by the orbital wind. No one knew, however, what molecules were involved, how they were formed, which atmospheric gases were important or what role, if any, the surface of the shuttle played in the light-emission process.

The first experiment to fly was designed to determine the identity of the molecules responsible for the glow by measuring the wavelength of light they emitted. The energy states any given molecule can occupy are determined by its atomic structure and makeup. When molecules move between states they absorb or emit light at specific wavelengths, corresponding to the energy difference between their initial and final state. Hence, the wavelengths of light that individual molecules emit or absorb can reveal their identity.

The first photographs of the shuttle glow had shown an orange color, but more detailed spectral information was necessary to identify the molecules responsible. Stephen B. Mende and his co-workers at the Lockheed Palo Alto Research Laboratories arranged for Thomas K. Mattingly II, the flight commander of shuttle flight STS-4, in June, 1982, to take pictures of the shuttle glow with a camera that had a diffraction grating over its lens. The grating broke up the light from the glow into its component colors (as a prism does) so that their intensities could be measured individually.

The results of this experiment were surprising. Mende found that the spectrum of the glow consisted of a continuous, unstructured emission that peaked in the yellow-red portion of the spectrum and extended into the infrared. Most atoms and molecules emit light at specific wavelengths; their spectrums consist of discrete lines. Among the very few molecules with emission spectrums that nearly match the observed spectrum are nitric oxide (NO) and nitrogen dioxide (NO_2). They are now considered the most likely source of shuttle glow.

The STS-4 flight provided another important clue to the glow's origin: the shuttle orbited at an altitude of 300 kilometers (compared with 240 kilometers for the mission during which the glow was first observed), and the glow observed during the later flight was significantly less intense. The atmospheric density was lower at the higher altitude, but more important, the estimated composition of the gases striking the shuttle was different as well. The intensity of the glow decreased at the higher altitude by about the same proportion as the concentration of atomic oxygen decreased: by a factor of about three. This correlation implicated atomic oxygen as an important factor in generating the molecules responsible for shuttle glow.

Experiments on board two subsequent flights probed the relation between the intensity of shuttle glow and the composition of the surface immediately below the glowing layer of gas. Samples of different materials were attached to the shuttle's manipulator arm, which was then stretched away from the shuttle so that the flow of atmospheric gases could strike the samples directly.

Different materials induced different intensities of glow; strangely, however, materials that caused the most intense glow were the ones that suffered the least surface erosion and mass loss. The brightest glow came from a black paint that is used to coat optical systems to reduce scattered light; it is highly resistant to attack by atomic oxygen. Conversely, materials that erode easily, such as polyethylene, produced little glow.

No material has been found so far that both produces a strong glow and erodes readily. This fact narrows the range of potential mechanisms for shuttle glow. If chemical reactions between surface materials and atomic oxygen were responsible for the compounds that produce the glow, erosion and glow would be positively, rather than negatively, correlated.

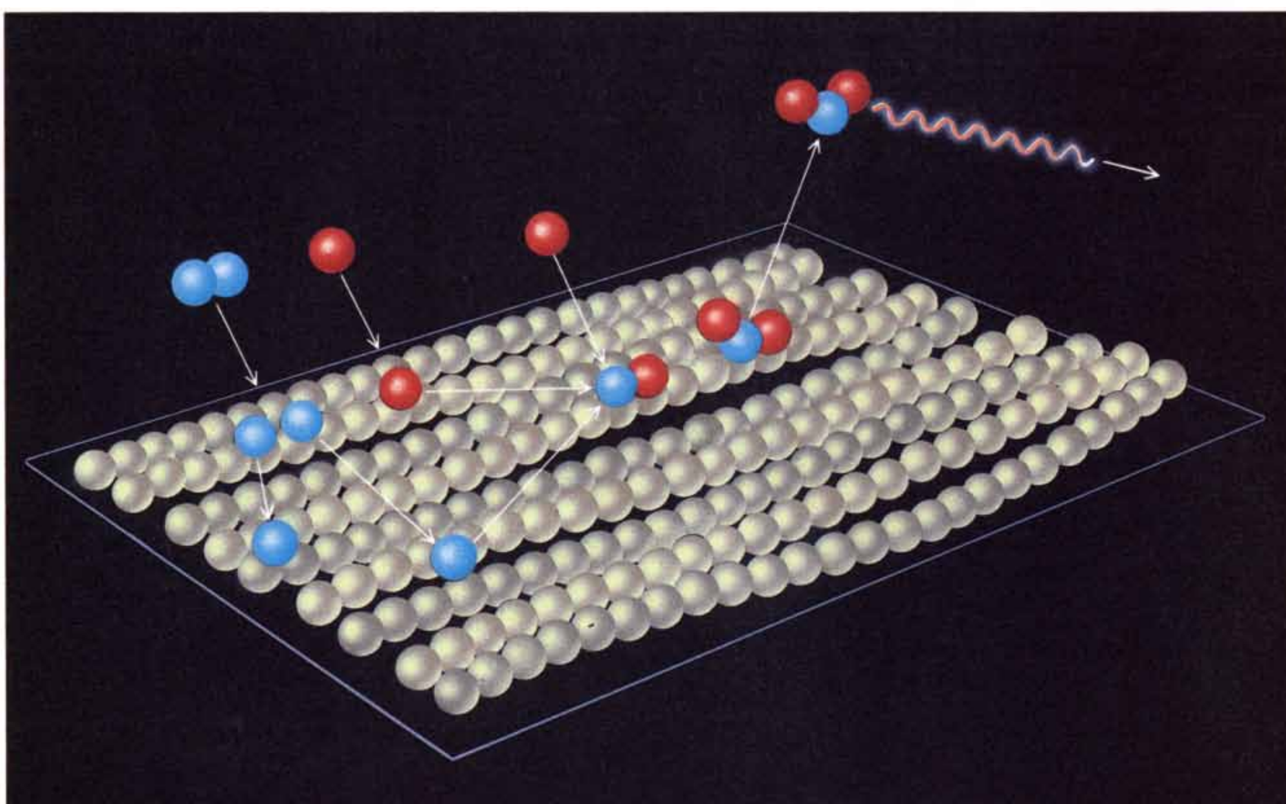
It is more likely that the surface acts as a catalyst, promoting reactions between atmospheric atoms and molecules to cause the glow. Catalysts enable chemical reactions to take place faster than they otherwise would, but the catalysts remain unchanged in the process. Variations in the intensity of the glow produced by different surfaces may reflect their varying efficiency as catalysts.

That efficiency might vary not only with the composition of a surface but also with its temperature. Mende has proposed that cold surfaces may promote the formation of the molecules that produce shuttle glow better than hot surfaces do. Impinging atoms and molecules stick to a cold surface longer and so have more time to combine into glow-producing forms. The black paint that produced an intense glow radiated heat more readily than the other surfaces tested, and so it cooled faster. Similarly, the insulating tiles on the shuttle's tail and engine pods are colder than other spacecraft surfaces. This could explain why the glow is more intense on the shuttle than on other spacecraft, which are built of different materials.

Given what is known so far, most workers have accepted in broad terms a three-step mechanism for the glow. Atoms and molecules from the upper atmosphere first collide with the spacecraft surface and become adsorbed. Next, as these species wander around on the surface, some of them meet and combine to form excited molecules. Finally, as excited molecules desorb from the surface, they relax into lower-energy states and emit light.

The inverse relation between glow and erosion implies that the surface material does not react directly with the atmospheric gases, and so the components of the glowing molecules must come from the orbital wind. The atmosphere at shuttle altitudes is composed mostly of atomic oxygen (80 percent) and molecular nitrogen (20 percent). Molecules made up of nitrogen and oxygen atoms are therefore the most likely candidates for producing the glow. Even though the laboratory emission spectrums of NO and NO_2 match the shuttle data fairly well, the experiments carried out on board the shuttle do not have the precision required to determine conclusively that these two molecules cause the glow.

Neither NO nor NO_2 is present in the atmosphere at shuttle altitudes in any appreciable concentration. The molecules must therefore be manufactured from nitrogen and oxygen atoms adsorbed on the glow-producing surface. In the first step of the proposed mechanism, then, atmospheric oxygen and nitrogen atoms collide with the surface and become adsorbed. The oxygen atoms come directly from the atmosphere, whereas the nitrogen atoms come from molecular nitrogen that dissociates on impact. (The relative velocity of the nitrogen molecules and the shuttle, about eight kilometers per second, corresponds to



SHUTTLE GLOW is a multistep process in which the spacecraft surface acts as a catalyst for chemical reactions but is itself unaffected. Atomic oxygen (*red*) and molecular nitrogen (*blue*) strike the surface (*left*); the nitrogen molecules dissociate owing to the kinetic energy of impact at orbital speeds. The atoms wander across the surface until they collide to form molecules of nitric oxide (NO) that are subsequently struck by incoming oxygen atoms to form nitrogen dioxide (NO_2). The energy released by the reaction causes the molecules to leave the surface in an excited state, and so they emit a photon of light.

an energy of 9.3 electron volts, comparable to the 9.8 eV of the bond holding the two atoms together in a nitrogen molecule.)

The oxygen and nitrogen atoms migrate across the surface until they meet and react to form NO. Two sequences of events are then possible. In the first, the energy released by the reaction excites the NO molecule to a high-energy state; if the molecule leaves the surface, it will emit this energy in the form of shuttle glow. In the second, the NO molecule is stabilized by transferring its energy to the surface. The molecule remains adsorbed until it is struck by an oxygen atom from the atmospheric wind to form NO_2. The energy from this reaction excites the NO_2 to an upper-energy level and kicks the molecule off the surface at high speed. The excited NO_2 molecule then emits light.

The fact that the glow is observed above the surface of the spacecraft would seem to contradict the idea that the glowing molecules are produced on the surface. Excited molecules that remain on the surface, however, can readily transfer their energy to the surface in the form of heat; they need not emit light. Molecules that desorb from the surface in an excited state must lose their energy by glowing.

The thickness of the glowing layer is related to the variations in the velocities of the excited molecules leaving the surface and in the lifetimes of their excited states. The distance each molecule travels before emitting a photon is equal to its velocity times the lifetime of the energy state it occupies. The excited state of NO_2 that yields emissions in the orange part of the spectrum has a lifetime of about 70 microseconds. If NO_2 is indeed the species responsible for glow, a layer 20 centimeters thick implies that molecules leave the shuttle surface at a velocity of about 2.5 kilometers per second. This velocity is much higher than could be accounted for by the average atmospheric temperature; instead about 25 percent of the kinetic energy of the incoming oxygen atom appears in the kinetic energy of the NO_2 molecule as it flies into space.

Experiments planned for future shuttle flights will probe both surface erosion and glow in greater detail. One important erosion experiment is the Long Duration Exposure Facility (LDEF), launched by the space shuttle *Challenger,* in April, 1984. The LDEF is a large cylinder covered with samples of spacecraft material; it also carries many experiments designed to measure the effects of the orbital environment on all parts of a spacecraft. The facility was originally meant to stay in orbit for only one year before being retrieved by the shuttle and returned to the ground for analysis, but schedule changes, followed by the *Challenger* disaster, left it in orbit.

Now, almost six years later, atmospheric drag has lowered the LDEF's orbit to the point that retrieval may not be possible much longer. The satellite may reenter the atmosphere and burn up (as did Skylab a decade ago) as soon as late January, 1990. It is with some sense of urgency that the shuttle *Columbia* is scheduled to rendezvous with the LDEF on December 21, following its launch of a communications satellite. If the LDEF is successfully brought back to the earth, it will provide a wealth of information about the effects of exposure to atomic oxygen fluxes as much as 10 times greater than that in other experiments.

A second erosion experiment is being undertaken jointly by the Lyndon B. Johnson Space Center, my research group at the Air Force's Geophysics Laboratory and several other NASA centers, along with international collaborators from Canada, the European Space Agency and Japan. This is the third in the series of experiments that started with the 1982 and 1983 flights. It will determine which materials react with atomic oxygen, how fast they react and how they can be protected from erosion.

The reaction rate for each material is proportional to the amount of material eroded divided by the number of oxygen atoms that strike the surface. Polymers that erode quickly lose about 3×10^{-24} cubic centimeter per incoming atom; that is, for each atom of polymer removed, about 10 to 100 oxygen atoms strike the surface.

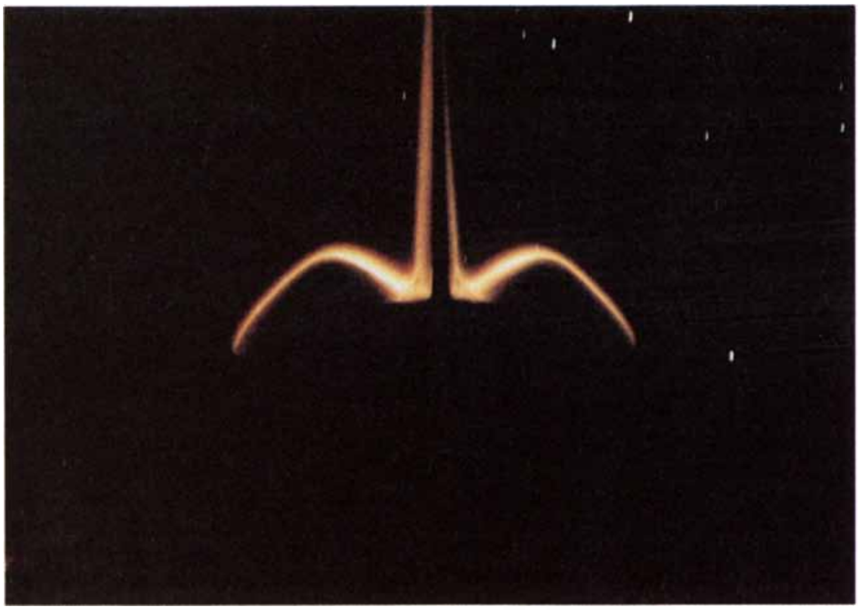
In earlier experiments carried out to investigate surface erosion, the amount of material eroded was measured after the flight, but the amount of oxygen striking each surface was not measured. Instead oxygen fluxes were calculated from computer models of the upper atmosphere. Although such models are accurate to within 20 percent, on the average, they do not provide accurate predictions of day-to-day fluctuations in oxygen concentration. On any given day the predictions of the model may be off by as much as a factor of two.

Our experiment, which is to fly in late 1990 or early 1991, will employ a mass spectrometer to measure directly the amount of oxygen striking test surfaces during the flight. The instrument, built at the Air Force's Geophysics Laboratory, can measure the composition of orbital gases to one part in 10,000 as they flow into it. It is a spaceflight veteran, having already measured the chemical properties of the shuttle environment during the craft's fourth flight, in June, 1982.

In addition to measuring the concentration of atomic oxygen, the mass spectrometer will also look directly at five sample surfaces mounted in a carousel in an attempt to detect and measure the products of any surface reactions. The materials that have tentatively been chosen are Kapton, carbon film, polyethylene, fluorinated polystyrene and either gold or anodized aluminum. The first three substances will contain carbon 13 so that their reaction products will be distinguishable from carbon compounds present in the ambient gases (which would contain the more common isotope carbon 12). Identifying the molecules generated in the reactions, either simple molecules such as water or carbon dioxide or more complex molecules derived from the structure of the samples, will help determine the chemical mechanism of surface erosion. The carousel will also carry one of the materials that has been found to glow strongly, such as anodized aluminum; we hope to identify the surface-reaction products from this sample and thereby confirm the identity of the glowing molecules.

Deeper understanding of the glow mechanism, however, will come from studying more detailed spectrums of the light. Planned glow experiments will obtain high-resolution spectrums from a variety of surfaces to gain a more precise picture of the levels and distribution of the energy states that produce the glow. One NASA experiment to be carried out by Leger's group at the Johnson Space Center will measure the intensity of the glow in the ultraviolet and infrared portions of the spectrum as well as in the visible range. Such spectral information will allow researchers to understand just how the glowing molecules are formed. The temperature of the surface causing the glow will be controlled to test the dependence of glow intensity on surface temperature.

Two other glow experiments are being planned by the Air Force's Geophysics Laboratory; the first is a set of infrared detectors, planned to fly in 1990, that will look straight out of the payload bay to detect glow along their line of sight. The shuttle will be maneuvered so as to change the orientation of the detectors with respect to



SHUTTLE TAIL AND ENGINE PODS glow in this picture taken from the aft flight deck of the craft during its eighth flight, in September, 1983. The engine pods were facing into the orbital wind during the exposure and so glow more strongly than the tail.

the atmospheric wind. The second experiment will be a more complete investigation along the lines of the NASA glow experiment.

Both glow and surface erosion pose serious problems for the design of future spacecraft. Scientific instruments such as telescopes obtain their information by gathering light, as do operational satellites such as the Landsat remote-sensing system. Any glow produced by an instrument itself or by the spacecraft it rides on could interfere with the phenomenon the instrument is designed to measure. At worst, glow could mask information completely. An ironic example of the problem is the black paint that has been used to coat the inside of optical instruments to reduce scattered light. This paint, which was at one time to coat the interior of the Hubble Space Telescope, produces a more intense glow than any other material tested thus far.

Erosion, for its part, strikes at the structural integrity of spacecraft; it raises the possibility that future space structures will have to be much thicker and heavier than now envisioned. Because of the cost of lifting objects to orbit (currently about $4,000 per pound), designers make them as light as possible. Objects that operate in the microgravity environment of space can be much lighter and weaker than they would have to be if they operated on the earth's surface. The robotic arm in the space shuttle's payload bay, for example, can lift satellites weighing thousands of kilograms out of the bay in orbit, but it cannot even support its own weight on the ground.

The erosion experiments carried out on board the shuttle, however, indicate that light, thin structures could lose a substantial fraction of their thickness and thus their strength during the course of long-term exposure to atomic oxygen. Erosion is particularly critical for the planned U.S. space station, which would remain in low earth orbit for at least 30 years.

Composites made of carbon fiber and epoxy have been proposed as the primary structural material for the station because of their high strength and light weight. During the course of the station's lifetime, however, carbon-fiber structural members could lose up to half of their intended .15-centimeter thickness to erosion by atomic oxygen. (The erosion rate anticipated for the space station is lower than that seen in shuttle experiments because the station will orbit at a much higher altitude.) The 13-micron-thick Kapton films designed to support flexible arrays of solar panels could disappear completely within a year. (Such erosion has not endangered the solar panels of communications satellites, because they orbit at still higher altitudes, where gas concentrations are negligible.)

The problems could be solved with thicker materials or by the replacement of structural members periodically during the life of the station—both costly solutions. A more attractive solution would be to develop new materials or protective coatings that can withstand bombardment by atomic oxygen. The experiments soon to fly on board the shuttle may point to appropriate materials. At the same time, coatings should be formulated so as not to produce glow that could interfere with optical observations.

Seven years ago no one knew that spacecraft could weather away or give off light as they circled through the earth's tenuous uppermost atmosphere. Now both phenomena are the subjects of intense research. Early experiments have tantalized researchers with hints about the surface chemistry that occurs in orbit. Detailed knowledge of such reactions will determine just how easily human beings and their equipment can function for extended periods in low earth orbit.

Glow and surface erosion went unrecognized until the development of a spacecraft that could fly back to the earth and be examined closely. That same spacecraft—the space shuttle—will now carry into orbit the experiments that will determine just what accounts for glow and erosion and whether effective measures can be taken to mitigate them.

FURTHER READING

OBSERVATIONS OF OPTICAL EMISSIONS ON STS-4. S. B. Mende, O. K. Garriott and P. M. Banks in *Geophysical Research Letters,* Vol. 10, No. 2, pages 122–125; February, 1983.

SPACE SHUTTLE GLOW OBSERVATIONS. P. M. Banks, P. R. Williamson and W. J. Raitt in *Geophysical Research Letters,* Vol. 10, No. 2, pages 118–121; February, 1983.

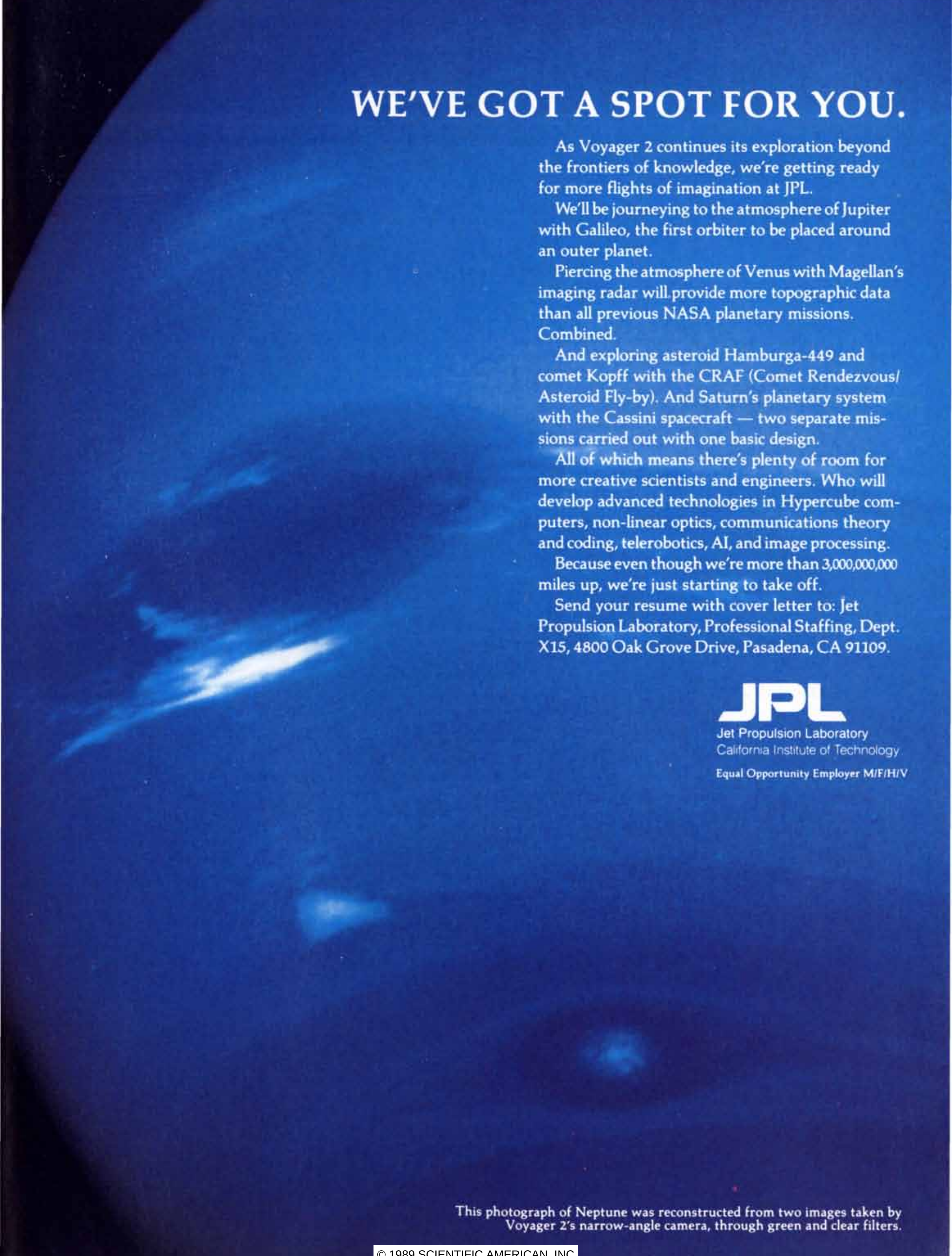
REVIEW OF THE VEHICLE GLOW. B. D. Green. American Institute of Aeronautics and Astronautics paper AIAA-85-6095-CP. AIAA Shuttle Environment and Operations II Conference, November 13–15, 1985.

SPACE VEHICLE GLOW MEASUREMENTS ON STS 41-D. S. B. Mende et al. in *Journal of Spacecraft and Rockets,* Vol. 23, No. 2, pages 189–193; March/April, 1986.

PROTECTING SPACECRAFT FROM ATOMIC OXYGEN. Lubert J. Leger and James T. Visentine in *Aerospace America,* Vol. 24, No. 7, pages 32–35; July, 1986.

THE EFFECT OF TEMPERATURE ON SHUTTLE SPACE GLOW. G. R. Swenson, S. B. Mende and E. J. Llewellyn in *Nature,* Vol. 323, No. 6088, pages 519–522; October 9, 1986.

ATOMIC OXYGEN EFFECTS MEASUREMENTS FOR SHUTTLE MISSIONS STS-8 AND 41-G. James T. Visentine. NASA Technical Memorandum 100459, Vols. 1 and 2. National Technical Information Service, September, 1988.



This photograph of Neptune was reconstructed from two images taken by Voyager 2's narrow-angle camera, through green and clear filters.

The World's Oldest Road

It is the Sweet Track of England: a 6,000-year-old wooden walkway discovered in a peat bog by a laborer named Raymond Sweet. The remarkably well-preserved wood reveals much about the track's builders

by John M. Coles

One day in 1970 a peat cutter named Raymond Sweet was busy at his typical winter occupation: clearing weeds from drainage ditches that vein the broad peat bogs of the Somerset Levels, flatlands in southwestern England. As he dug, his spade hit a hard object. Peat is dead plant matter and is soft, and so resistance meant that some substance other than peat, such as a root, was present. This particular object was not a root, but a plank of hard wood.

Sweet suspected that the wood was old, for it was buried near the bottom of what once had been many meters of peat, laid down over thousands of years. He delivered a fragment to the peat company, the Eclipse Peat Works, which in turn sent it to me at the University of Cambridge. I had been doing archaeological work in the Levels for several years, and my interest in old wood was well known among the peat cutters.

A prompt visit to the site, about 250 miles from Cambridge, confirmed that the plank was indeed old: it had been worked by sharp tools of stone and wood, raising the exciting possibility that it was a relic of a Neolithic (late Stone Age) community. Studies I began the following summer with a team of students soon revealed that the plank was but one piece among many thousands forming an 1,800-meter-long road through a swamp. Radiocarbon analyses indicated that the wood was extraordinarily old: it dated from some 6,000 years ago, early in the Neolithic period. Sweet had found the oldest road ever discovered.

Our investigations of the road, now known as the Sweet Track, continued for a decade, during which we traced its route and rapidly excavated 400 meters of it in areas where commercial peat harvesting was about to destroy the remains of the track. We also analyzed the construction of the road and gathered other clues to the skills and activities of the track builders. The peat company cooperated wonderfully, but at times we were nearly thwarted by other enterprises also engaged in drainage. Once we happened on a threatened site in the nick of time: a machine had just begun to gouge away the track. Had we arrived just 15 minutes later, we would have found only a heap of shattered wood.

The decade of investigation yielded rich clues to the lives of the Sweet Track's builders and revealed technological advances that were previously unknown for Neolithic times. To a great extent our discoveries were made possible by the abundance and excellent state of preservation of the wood in and alongside the ancient road. Wood was central in early technologies, but it is a rare find in Neolithic sites, most of which are dry. Being organic, wood dries out and decomposes rather quickly when it is exposed to air. It survived quite well in the Levels because the acidic, waterlogged peat kept the wood moist, inhibited the growth of bacteria and fungi that would have broken it down, and protected it from discovery by animals and people.

Given that the peat bogs account for both the preservation and the eventual discovery of the Sweet Track, it is perhaps fitting that the details of what we found be preceded by a history of the formation and exploitation of peat in the Levels.

The Levels of today can be thought of as a broad trough, roughly bounded on the north and south by the high Mendip and Quantock hills, on the west by the Bristol Channel and on the east by the gently rising hinterland of Somerset County. (The Levels lie about 50 miles west of Stonehenge, whose earliest arches were erected about 2,000 years after the Sweet Track.)

For generations the Levels have been a land of green meadows, black peat bogs and water-filled drainage ditches. Once, though, the land was an ancient fjord. It had been flooded when the ice sheets of the Northern Hemisphere began to melt roughly 10,000 years ago, raising the sea level. At that time the surface of the water was broken only by occasional small islands of sand or rock and by the Polden Hills, which separate the northern part of the Levels—the Brue River valley—from the southern part, a region called Sedgemoor. The basin has been more or less wet ever since, although not always because of the sea.

By 5000 B.C. waters from the Atlantic Ocean had blanketed the valley floor with thick, salty sediments of clay. Within 500 years the sea was virtually excluded, probably by sandbanks that formed at the western border of the Levels and by sediments deposited there by such rivers as the Brue, the Axe and the Parrett. The valley became a brackish swamp as high rainfall on the surrounding hills combined with shallow gradients in the valley to cause the rivers and the meandering streams of the Levels to overflow repeatedly. Beds of reeds took root in the marine clays, forming a freshwater "wetscape."

The permanent waterlogging initiated peat formation, as the reeds grew, died and fell into the swamp. By 4000 B.C. the layer of reed peat was perhaps a meter thick. For the moment, the valley was barely touched by human

JOHN M. COLES, a fellow of Fitzwilliam College at the University of Cambridge, retired from Cambridge as professor of European prehistory in 1986. He earned a Ph.D. in archaeology from the University of Edinburgh in 1959 and was awarded a doctorate of science from Cambridge in 1980, a degree acknowledging his many scientific publications. Since leaving Cambridge he has directed various projects and worked as an archaeological consultant. Coles specializes in wetland and experimental archaeology and is currently studying Scandinavian rock art.

beings; it was occupied instead by wild game, fish and a multitude of invertebrate animals.

Century after waterlogged century, the peat continued to accumulate, ultimately reaching a thickness of about 10 meters. As it formed, layer on layer, it trapped pollen. Pollen grains, which can travel some distance before settling, can be readily distinguished by their shape and counted under the microscope. Hence, the trapped pollen grains provide a record of natural vegetation and sometimes of cultivated crops (reflecting human activity) over a broad region.

In about 4000 B.C. the first farmers of Britain began arriving from the European mainland, seeking lands to clear and cultivate. No better place could have been found than the Levels. Settlers could turn to the low-lying wetland for its wild resources of food and material (such as reed for thatch). They could search the edges of the swamp for willow and alder (valuable for making rope, stakes and woven panels called hurdles) and for fish, beaver, otter, deer and wild pigs. And they could clear the forests of the higher, drier slopes to obtain timber and to establish croplands and grazing areas.

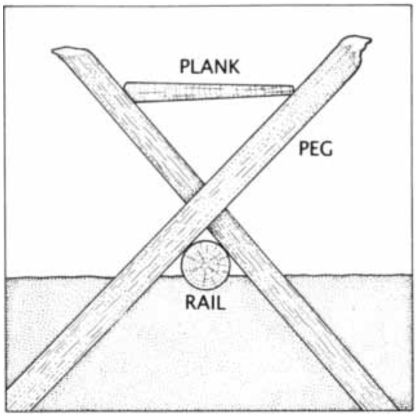
The pollen record shows that episodes of forest clearing and subsequent cultivation continued on the hills for several thousand years. The episodic disappearance of cereal and pasture-weed pollen, along with the return of pollen from depleted tree species, marks the periodic abandonment of settlements here and there.

The pollen record ends in about A.D. 400 when peat growth ceased. Massive floods swept across the Levels, preventing much new plant growth. Later the area dried out somewhat, but by then conditions were no longer suitable for peat formation.

The peat was left unharvested by human beings for some time, but it has been cut for at least the last 1,000 years. (Initially it was burned for fuel; more recently it has mainly been used as fertilizer.) Indeed, medieval records recount disputes between the great abbeys of Glastonbury and Wells (at the eastern edge of the Levels) over control of the peat as well as over rights to other resources such as summer pastures and reed beds.

Because the Levels have a high water

SWEET TRACK, which once traversed 1,800 meters (1.1 miles) of a reed-filled swamp, looked something like the artist's representation at the left. A traveler on the road would have carried tools of stone and wood and worn clothing made from animal skins or plant fibers. The road succumbed (perhaps within a decade or two after it was built) to rising water and accumulating peat. Indeed, the water apparently dislodged many of the planks that made up the walkway's surface. Hence, when parts of the track were excavated (*right*), the author found few such planks still in place, although much of the roundwood understructure was basically intact. One plank that did stay put is visible in the center of the photograph.



CROSS-SECTIONAL VIEW of the Sweet Track reveals its substructure, which consisted of single rails laid end to end and long, sharpened pegs. The pegs were hammered into the swamp so that each pair (or grouping) became an X whose vertex rested on the rail. The brackets thus formed by the pegs held the planks.

table, until recently only the layers of peat close to the surface were cut. In the past several decades, however, electric and diesel pumps have transformed the peat-cutting industry. Now the whole of the Levels is being drained. Large-scale peat cutting has uncovered hundreds of artifacts from the distant past, including many ancient roads, but the peat-cutting machines have also obliterated much of the prehistoric record. Slicing their way across the peat fields, they have lowered the surface year by year, until in places nothing is left to cover the underlying marine clays. Fields that once supported reeds, cotton grass, rushes and bright mosses but that are now flat, gray and devoid of life are a sorry sight.

The place where the Sweet Track was found—Shapwick Heath—was just such a desolate place in 1970 when Raymond Sweet was working there. Almost all of the peat had been cut away, leaving only one meter of the lowest, oldest material awaiting removal. The area, which is just north of the Polden Hills in the southern half of the Brue River valley, contrasted sharply with the buzzing reed-filled swamp it had been 6,000 years earlier, when the track was built.

Why was the track constructed, and what has been learned from it? The evidence indicates that the Sweet Track builders had a settlement on the gentle slopes of the Polden Hills and built the road to serve as a footpath linking two islands on which the settlers were also active: a protruding island of sand just north of the Polden Hills and Westhay Island, a relatively large outcrop of limestone about two kilometers north of the sand island. The track builders gathered timber from the southern island, just as they did from the hills, and they also cultivated land there. How Westhay was exploited is less clear. It too supported woodland and was a source of timber; parts may have been made into cropland or pastures for domestic animals, perhaps sheep and cattle. Certain evidence also suggests that Westhay was settled; some of the track builders may have lived there rather than on the Polden Hills.

To construct the road, the workers selected about a dozen species of wood from the forests both north and south of the planned track. They felled the larger trees (mainly oaks, ashes, elms and limes, or lindens), some of which were a meter in diameter, to make long planks for the surface. The trees were debarked, cut and split in the forest. Roundwood (branches and young trees) of many species—such as hazel, ash, alder and elm—was collected for the road's underlying support structure, which consisted of rails (long, fairly thick pieces of roundwood) and pegs (thinner pieces that were sharpened at one end).

Once the wood had been brought to the northern and southern terminals, construction began. First, single rails were set out end to end on the surface of the swamp. Next, pairs of pegs were hammered (points down) into the peat so that they formed a tall X whose vertex rested on a rail. The X's were spaced a meter or so apart, and their upper arms formed a V-shaped scaffold into which a plank could be set. Planks were laid end to end and parallel to the rails (perhaps after being notched for a better fit with the pegs), forming a narrow walkway elevated about 40 centimeters above the rails [*see illustration on this page*]. (The rails were critical because without them the weight of the planks would have caused the track to subside gradually into the peat. The rails served to spread the load and thereby prevent or delay such sinking.)

Experiments in building a replica of the Sweet Track indicate that the components could be assembled quickly and easily to make a firm walkway raised about half of a meter above the swamp—significantly above normal water levels and high enough to provide a fairly dry surface for pedestrians. Indeed, the ease with which the rails, pegs and planks could be put in place suggested to us that the entire 1.1-mile-long track could have been assembled by about 10 men in a single day! This is not in any way to denigrate the achievement of the track builders but rather to emphasize their skill in woodworking and in engineering such a complex structure.

The evidence for this scenario of collaboration and rapid construction comes in part from an extensive analysis of the tree rings in the planks, rails and pegs. Each ring represents a year of growth, and the width of the ring reflects the year's growing conditions. The presence of a closely similar sequence of thick and thin rings in a number of wood samples indicates that the trees from which the samples came grew in the same region, under the same conditions and at the same time.

Our tree-ring analyses have demonstrated that almost every piece of oak and ash (which account for most of the planks) came from trees felled in the same year—at the same point, that is, in a sequence of matching rings. Hazels and other trees cut for pegs had similarly been felled in a single year, presumably at the same time as the planks were cut.

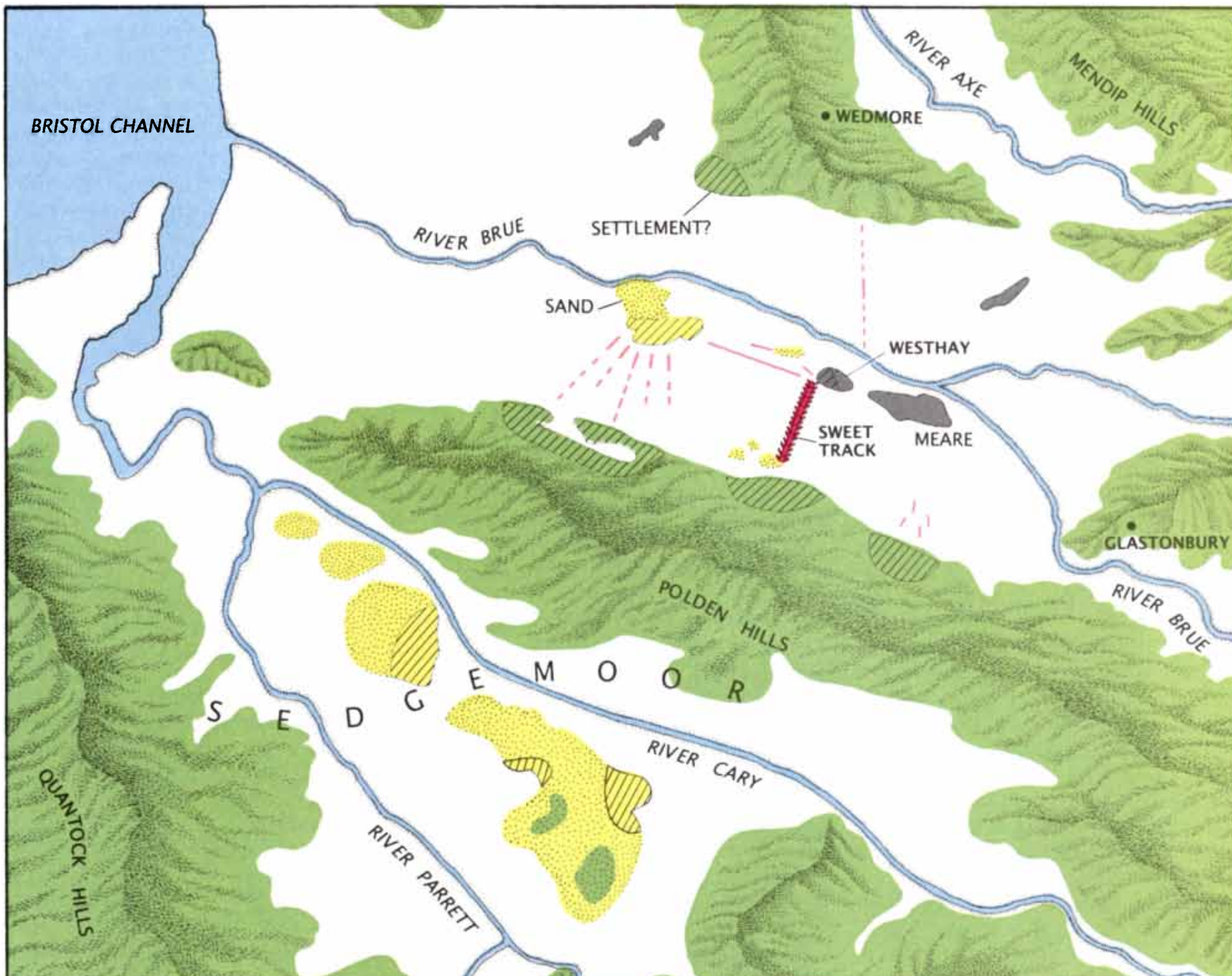
The probability that the difficult undertaking of building the track was completed in a short time is suggestive of strong social organization. Apparently, the settlers agreed that a track was needed to connect their settlements and fields; they concurred on its precise position, and they dispatched surveyors to lay out the track line. (We know that a preliminary route was established in advance of construction because we recovered wood posts that roughly outlined the course of the track but had no structural role in it.) The settlers also had to organize the labor needed for the major tasks: felling the trees, preparing the required track components and transporting the stockpiles down to the swamp. (The construction itself was probably the least difficult task.)

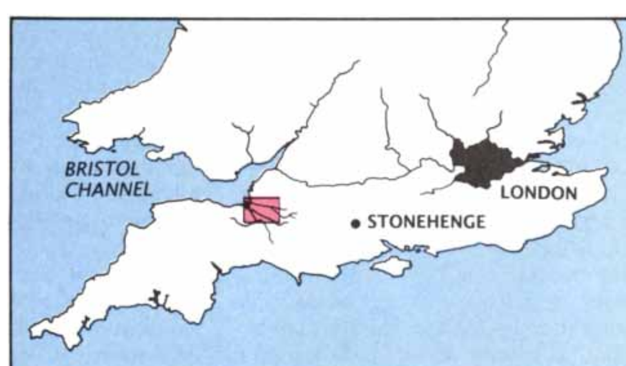
Analyses of the tree rings also provide some evidence that the builders of the Sweet Track were not the first inhabitants of the Levels. The wood in the northern part of the track indicates that the northern woodland included 400-year-old oaks, whose maturity suggests that the forest had been undisturbed until the track was built. In contrast, the wood from the southern part of the track indicates that the southern woodland supported 120-year-old oaks, which means the oak stands there had been cleared about a century earlier, by a group of Britain's earliest farmers—probably the track builders' direct ancestors.

The tree-ring data suggest as well that the track served its purpose for just a short time. Only a few planks came from trees felled later than that of the bulk of the wood, and the dates extend to only 11 years after the building of the track. The timber cut in the course of that short time was inserted for repairs in places where seasonal floods carried away planks or otherwise damaged the track.

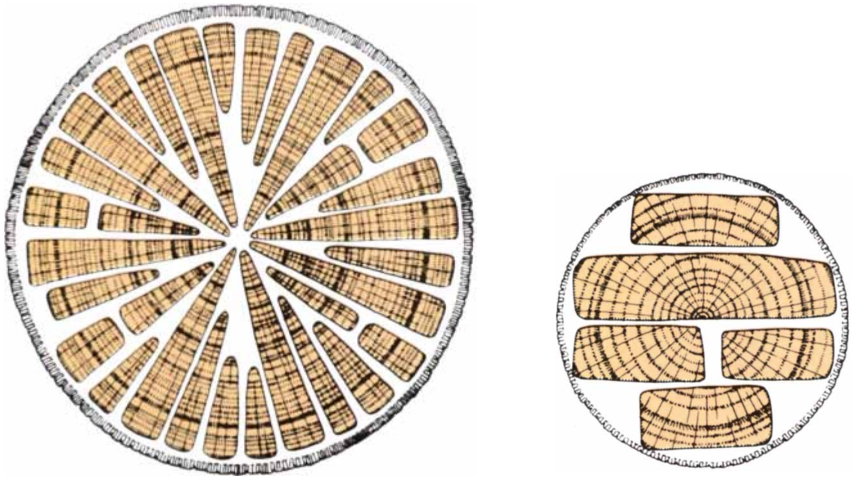
The absence of still later repairs suggests that repeated flooding, the relentless growth of reeds, the accumulation of peat beneath and around the track and the consequent rising of the water level eventually doomed it. The walkway was no longer high enough above the swamp to be negotiable or repairable. And so we think that a child who may have helped to build the track would barely have been an adult when the road was finally abandoned and left for the reeds and peats to bury and hide.

The quality of the woodwork in the Sweet Track reveals something more about the people who made the road and the way they went about it. Our experiments—which involved felling oak and ash with original Neolithic tools, splitting trunks into planks and axing notches and sometimes also holes into the resulting boards—show that the track builders had an intimate knowledge of the properties of wood. They also wielded their wood and stone tools with remarkable skill. (Ne-





SOMERSET LEVELS, where the Sweet Track was found, are in southwestern England. They consist of 100,000 acres of low-lying land (*white*) that is bounded on the west by the Bristol Channel and on all other sides by hills; the green shading represents elevations of 15 meters or more above sea level. The Levels are dotted with raised "islands" of sand (*stipple*) and rock (*gray*) and are partially divided by a limestone range called the Polden Hills. In addition to the Sweet Track (*red*), which was built around 4000 B.C. to connect a small sand island with Westhay Island, there are several younger walkways in the Levels, dated to between 2500 and 3500 B.C.; the excavated parts of these tracks (*pink lines*) and their probable courses (*broken lines*) are marked. Areas where settlements were probably established between 2500 and 4000 B.C. are also shown (*hatching*). The river courses indicated are present day.



OAK TIMBERS, a major source of planks in the Sweet Track, were fashioned into long boards in two ways, depending on the size of the felled trees. Tall 400-year-old oaks with diameters as large as a meter were taken from a forest north of the track. Cleverly, the builders split these radially (*left*) to form planks that were from 25 to 40 centimeters wide. Younger trees, from south of the track, were about half of a meter in diameter—too small to be cut radially—and so they were split tangentially (*right*).

olithic farmers had the benefit of neither metal axes nor saws.)

In order to fashion oak planks, for example, they felled trees that would yield strong, straight timbers, free of knots and side branches. Then, exploiting the tendency of oak logs to crack radially, they split the larger felled trunks into planks by hammering wedges into the rays [*see illustration above*]. Relatively young oaks cut in that way would have yielded planks that were too narrow, and so these were split laterally, a more difficult task.

The sharpened ends of the pegs further testify to the builders' skills. The workers possessed stone axes of high quality, resharpened them frequently and knew the differing properties of various roundwoods. For example, hazel and alder were chopped with short strokes, whereas willow and poplar, which are difficult to cut in that way, were shaved into long points.

The pegs (and there were 6,000 of them) further taught us that the track builders were engaged in woodland management—the earliest known in the world. They selectively felled hazel and ash and probably oak and alder, too, in order to encourage the rapid growth of many shoots from the remaining stumps. The practice, called coppicing, yields long, straight rodlike shoots that can be cut after a few years. We have many pegs whose tree rings and shape show the characteristic quick growth and straightness of a managed, coppiced woodland.

No investigation of a wet archaeological site can succeed without extensive, multidisciplinary environmental sampling—analyses not only of the pollen but also of the macroscopic remains of plants and animals. For example, the beetles whose chitinous exoskeletons were partly preserved in the Levels have enabled us to give specificity to what had previously been broad generalizations about the Neolithic climate in southern Britain. Several species of beetle can still be found in the Levels. Others have disappeared, in some instances because of their intolerance of climatic change. One of the latter, *Oodes gracilis*, can be found today only in parts of continental Europe that have cooler winters and warmer summers than the Levels do. The beetle's presence thus indicates that winter temperatures in southern Britain at the time of the Sweet Track's construction were from two to four degrees Celsius colder than they are today and that summer temperatures were from two to three degrees warmer.

The pollen record in the Levels yielded other invaluable information. Pollen grains from the cereals and grasses that grow on open ground are our best evidence of farming on the hills and islands. Moreover, the range of pollen species we found confirms the existence of woodlands around the swamp. The pollen data, when combined with our knowledge of the wood in the track, made it possible to determine the composition of the upland forests. The general pattern was a mixed forest in which oaks, elms, ashes and limes were the common large trees, towering over an understory of hazels and hollies. Lime trees were once thought to be rare in the Neolithic period, but the pollen data and the track itself indicate that they flourished in the Levels, particularly on the Polden Hills.

A fine example of what can be learned from multidisciplinary studies is our finding that the middle portion of the track crossed open water. Here, instead of a single rail and neat sets of pegs, three rails were piled atop one another to gain extra height for the walkway, suggesting that the builders had to adapt their structural plan to cope with high water. Ultimately, they were unsuccessful: we found that here the planks of the walkway were in disarray, off to the side of the understructure, which indicates they had been washed away. Analyses of plant and animal remains confirm the presence of open water. Water lilies flourished along with the usual rushes and reeds; so did whirligig beetles, which live on the surface of water, and large raft spiders, which live on or near permanent pools of water, where they hunt aquatic insects and even small fish. The survival of remains of these soft-bodied spiders in the peat for 6,000 years is truly remarkable.

The artifacts left by those who walked the track offer further glimpses of activities in the swamp and beyond. The track was only about 30 centimeters wide, stretching through high reeds that must have masked the traveler's view of the farther terminal. View-obscuring reeds, high water and projecting pegs all combined to make the walkway precarious; one feature it lacked was a passing lane! Hence, one might expect to find items that were lost or broken along the track as well as residues of swampland activities.

Among our many finds were flakes. Such tools, which were produced by striking cores of fine, dark flint with a hammer made of stone or bone, are ubiquitous on Neolithic dryland sites. These samples interested us because we knew without a doubt that they had been undisturbed virtually from the moment of loss.

Several flake tools were in immaculate condition for "use-wear" analyses. With the aid of a microscope, one can determine what a tool was used for by examining the pattern of wear. On this basis, the flakes could be sorted into distinct groups. One was for cutting wood, one for reeds, one for other plants (which as yet are unidentified but left a mottled smear on the flakes) and one (represented by a single flake) for hide. The collected flakes are evidence that, as expected, the people hunted in the swamps and made use of reeds and other plants.

Flint arrowheads too were found, providing further evidence for hunting if not for warfare. Traces of glue, arrow shafts and binding string were preserved on some of the arrowheads—a rare find. Many fragments of bows were also recovered.

All the flints had been imported to the Levels, and so their presence points to contact with people outside the immediate region. Also imported were a flint axhead (from eastern England) and one, astonishingly, of jade (probably from the European alpine area), both of them in mint condition. We can state with assurance that these could not have had wood ax handles, because such handles would surely have been preserved by the peat. The axes' unhafted, unused condition suggests that they were perhaps being taken to a work site when their unlucky owners lost them in the course of what turned out to be a very expensive walk. It is also possible that the axes were deliberately deposited in the Levels for a purpose unknown to us, possibly as a dedication for the Sweet Track.

WORKERS excavating the Sweet Track had to stay off the surface of the site, and so they could often be found lying stomach-down on boards elevated over the road. Because the wood in the track and the peat in which the track is embedded are both quite soft, even a careful step could have easily crushed any hidden wood.

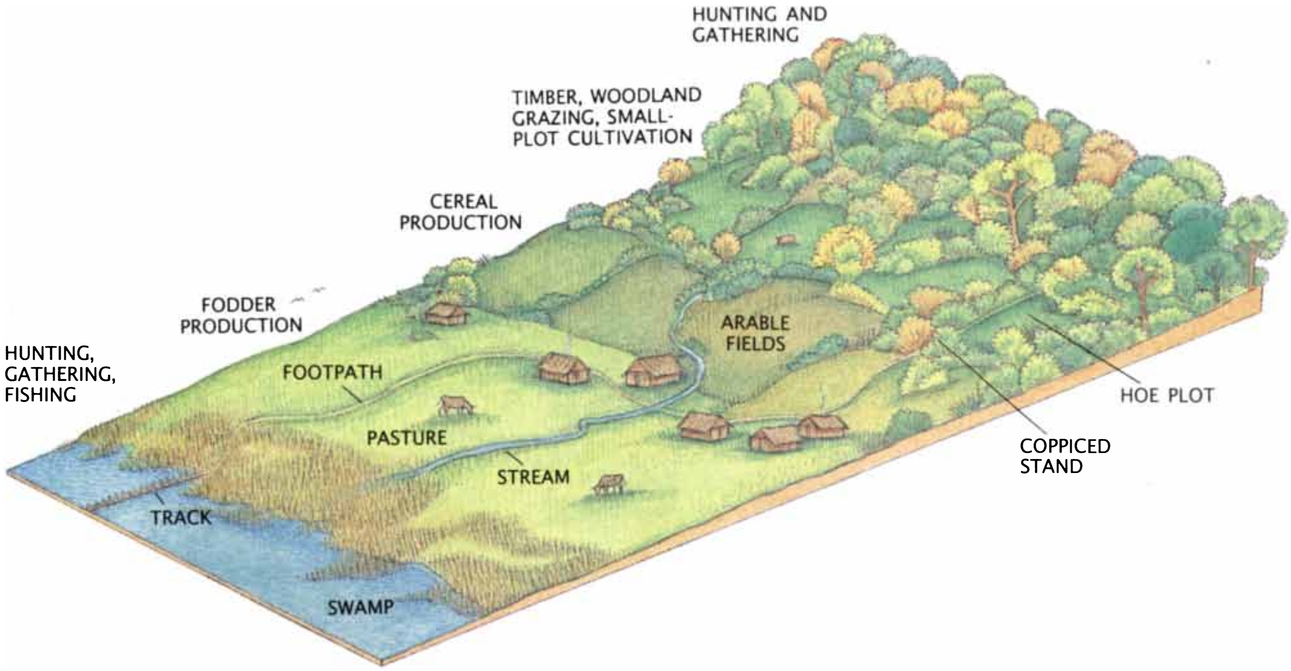
We also found pottery and various wood artifacts along the track. Some of the artifacts, such as solitary potsherds, may have been washed into the swamp from nearby settlements rather than deposited there directly. However, heaps of potsherds that form almost complete pots were probably dropped by travelers on the track itself. In one instance, the broken vessel held hazelnuts. In another instance a wood stirrer was being carried in the pot, which may have held some sort of gruel, although no traces of it survive.

The wood artifacts include digging tools (such as mattocks and spades), paddles, wedges, handles, a spoon, a comb, fasteners and a carved bowl. None of these are in any way spectacular in design or construction; they are the common variety of items one assumes existed. They are of value, however, because their likes have perished from all dryland sites of the same period.

One discovery was a total surprise: carefully smoothed pins of yew, whose function is unknown. They are thin bits of wood, approximately 20 centimeters long, curved to a point. They may have been decorative nose pins or, perhaps, clothes fasteners. If the latter, the pins are all that remain of the clothes and the people who wore them: the acidic peats have dissolved all skin and bone.

The loss of most nonplant material is the price paid for the exceptional survival of wood and other plant remains. Indeed, of the particular animals hunted or otherwise exploited by the track builders we know nothing concrete. We know there were rodents around, because teeth marks left on

SHARPENED END of a peg, shown shortly after it was uncovered, has clean facets, which is one of many indications that the builders of the Sweet Track had high-quality tools and kept them well sharpened. This peg is made of hazel; others were made of holly, alder, birch, apple, willow, ash, lime, elm and oak. Ancient pegs, rails and planks dry out rapidly when exposed to air. To prevent such destruction, most of the wood recovered from the track (including this piece) was preserved by replacing its water with a wax (polyethylene glycol), a process that takes about nine months.



VARIED ACTIVITIES of the track builders have been deduced from analyses of the wood in the track and from studies of pollen, artifacts and other materials buried in the peat surrounding the road. In the swamp the settlers hunted and fished and gathered such items as reeds for thatch. On higher ground, where they built their settlements, they cleared parts of the forest to obtain wood and to establish pastures and croplands; they also hunted in the undisturbed regions of the forest. They practiced coppicing as well: the cutting back of trees to encourage the rapid growth of rodlike shoots from the resulting stumps. Coppiced wood found in the track is the earliest evidence of woodland management yet to be discovered.

a couple of hazelnuts were made by a bank vole and a dormouse. These small animals, though, were probably of little interest to the farmers and hunters of the Levels 6,000 years ago.

To some people the extensive analysis of a short footpath through a swamp may seem an exercise of little merit. Yet the discoveries we have made reveal the importance of such work and of searching for other wetland sites.

Think of what we have learned about the past from our treasure trove of wood and our environmental studies. We know the vegetation of the swamp and ancient woodland and many of the activities carried out both in the swamp and on the hills. We know the temperature range of the time. In the coppiced pegs and rails we have the first evidence of woodland management in the Neolithic period. The track also tells of the extraordinary woodmanship and carpentry of the builders, who had only fire, stone axe and wedge to handle great trees. Presumably, these were not the only people at the time who had such skills.

The track's quick burial, its structure and, above all, the tree-ring studies assure us that everything found on, in or beside the track was contemporaneous. This is not standard archaeological contemporaneity, which deals in spans of 200 or, with luck, 100 years. It is real time. Everything—artifacts of stone and pottery, wood planks and pegs, the swamp, the bank vole and dormouse and the dryland forests, fields and coppices—was in existence during a single span of time lasting less than a lifetime.

My story does not end here. Much of the track remains unexplored and will need constant attention if it is to be protected. Ancient structures buried in wetlands can deteriorate as the wet peats or silts are drained; they can decay even without being fully exposed to air. Our Sweet Track excavations were restricted to parts immediately threatened by peat cutting. About 900 of the remaining 1,400 meters lies beneath the peat of pastures that are slowly drying out. Negotiations are under way to ensure a halt to this desiccation. Fortunately, the remaining 500 meters has been acquired by conservation agencies from the peat company, which was willing to give up its rights to the area because of the importance of the buried track. This area has now been preserved by an ingenious system of clay banking around the reserve and pipes that bring in water to flood the peats every day in the dry season.

Continued preservation is crucial because the track and the materials buried beside it hold more clues to the track builders than can be gleaned by current technologies. For example, we think that food residues, textile fragments, fingerprints, perhaps blood and the remains of small invertebrates may still survive, along with other elusive clues to how the track builders lived and what challenges they faced. If the track can be protected, archaeologists will be able to return to it in the future, when they have the sophisticated tools and analytic methods needed to uncover more information. It is from wetlands such as the Somerset Levels that we will obtain evidence of the quality and precision needed to come close to understanding the dynamics of human behavior—which is, ultimately, the aim of archaeology.

FURTHER READING

PREHISTORY OF THE SOMERSET LEVELS. J. M. Coles and B. J. Orme. Somerset Levels Project, Universities of Cambridge and Exeter, 1982.

SOMERSET LEVELS PAPERS 10. Somerset Levels Project, Universities of Cambridge and Exeter, 1984.

SWEET TRACK TO GLASTONBURY: THE SOMERSET LEVELS IN PREHISTORY. Bryony and John Coles. Thames and Hudson, 1986.

Over 150 other airlines admire our maintenance so much, we work on their planes too.

How well an airline maintains its reputation depends largely on how well that airline maintains its planes. Which is why, after considering Lufthansa's record for on-time performance, so many other airlines have asked Lufthansa to help keep their planes in perfect working order.

After all, if so many other airlines ask for Lufthansa, shouldn't you?

People expect the world of us.

Lufthansa
German Airlines

Lufthansa is a participant in the mileage programs of United, Delta, USAir and Continental/Eastern. See your Travel Agent for details.

The Case for Methanol

The authors maintain that a move to pure methanol fuel would reduce vehicular emissions of hydrocarbons and greenhouse gases and could lessen U.S. dependence on foreign energy sources

by Charles L. Gray, Jr., and Jeffrey A. Alson

The private automobile has shaped U.S. society to a degree unparalleled by any other product of the industrial age. By providing mobility and convenience particularly attuned to the American desire for personal freedom, the automobile has come to dominate not only the nation's transportation network but also its very culture. And the automotive industry has become a pillar of the economy, accounting for more than 10 percent of the gross national product and some 20 percent of all consumer expenditures. Yet the automobile also threatens the quality of life, contaminating both urban air and the global atmosphere, where automobile emissions contribute to the greenhouse effect. The automotive industry must overcome unprecedented technical, political and social challenges if these serious environmental problems are to be solved.

To achieve this goal, we believe the nation must begin making a transition to a new automotive fuel. Having studied a wide range of alternatives, we think that fuel should be methanol. A move to methanol could achieve emission reductions far beyond those that are feasible even with advanced emission controls on gasoline vehicles. Although the past 15 years have seen substantial reductions in noxious pollutants and greenhouse gases from individual vehicles, the number of vehicles has been steadily increasing. Consequently, more than 100 cities still have ambient levels of carbon monoxide, particulate matter and ozone (generated from photochemical reactions with hydrocarbons from vehicle exhaust) that exceed the levels established by the Environmental Protection Agency to protect public health. As the nation's fleet continues to grow in the next decade, air quality will worsen unless vehicles can be developed that are much cleaner than those on the roads today.

CHARLES L. GRAY, JR., and JEFFREY A. ALSON work at the U.S. Environmental Protection Agency's Emission Control Technology Division in Ann Arbor, Mich., and have played key roles in the EPA's alternative-fuels program. Gray is director of the division, which is responsible for developing federal vehicle-emissions standards. Gray got his B.S. in chemical engineering from the University of Mississippi and his M.S. from the University of Michigan. Alson, who is assistant to Gray, has a B.S. in engineering from Purdue University.

Introducing methanol to the U.S. transportation infrastructure would require relatively modest changes for the automotive and energy industries. Our research has convinced us that this is the only practical means to achieve major reductions in vehicle emissions while maintaining the personal mobility that Americans have come to expect. Although there will be costs in making such a transition, there will also be significant benefits not only for the environment but most likely for the nation's economic health as well. We have incontrovertible evidence from vehicle tests and computer simulations that vehicles operating on pure methanol would bring about dramatic decreases in urban levels of ozone and toxic substances. What is more, methanol can be produced with current technologies from a variety of abundant sources, including natural gas, coal, wood and even organic garbage. By beginning a transition to methanol, the nation could ultimately lessen its dependence on foreign sources of energy.

The quest for cleaner automotive fuel has become particularly urgent. California and the northeastern states are calling for more stringent restrictions on gasoline-vehicle emissions than are currently in place at the federal level, and Congress is considering new amendments to the Clean Air Act. There is a growing realization that—short of a massive shift away from the private automobile—the only truly effective way to achieve significant further reductions in vehicle emissions is to replace conventional gasoline and diesel fuel with cleaner-burning fuels. Candidates include compressed natural gas, liquefied petroleum gas, electricity, ethanol ("grain alcohol") and methanol ("wood alcohol"). Synthetic gasolines are costly and offer no environmental advantages; they are no longer considered seriously.

The most promising alternatives are mainly carbon-based fuels whose molecules are smaller and simpler than those of gasoline. These molecules burn more cleanly than gasoline in part because they have no (or few) carbon-carbon bonds, and the hydrocarbons they do emit are less likely to generate ozone. The combustion of larger molecules, which have multiple carbon-carbon bonds, involves a more complex series of reactions, which increase the probability of incomplete combustion. On the practical side, the alternative fuels have drawbacks. Compressed natural gas must be kept under a pressure of from 2,000 to 3,000 pounds per square inch and so requires a set of heavy tanks—a serious liability in terms of a vehicle's performance and fuel efficiency. Liquefied petroleum gas faces fundamental limits on supply. Electric vehicles are now limited by battery technology: the batteries are heavy and need to be recharged frequently, a time-consuming process. Barring major technological breakthroughs, each of these energy sources will probably satisfy only certain market niches, for example, delivery trucks and buses.

Liquid fuels such as ethanol or methanol have important advantages: they have a higher energy content per volume than the other alternative fuels and would require minimal changes in the existing network for distributing motor fuel. Ethanol is commonly used as a gasoline supplement, but it is currently about twice as expensive as methanol, whose low cost is one of its most attractive features. Studies of the economics of producing and distributing methanol conclude that the fuel equivalent to a gallon of gasoline—about two gallons of methanol—would cost the consumer about $1.10. Methanol also has only half the vapor pressure of gasoline, so that evaporation will be a negligible source of pollution. Most important, methanol can reduce by 90 percent the vehicle emissions that form ground-level ozone, the most serious urban air pollutant. Finally, the properties of methanol make it possible to design a far more efficient automobile than those now on the road while still maintaining very low emission levels.

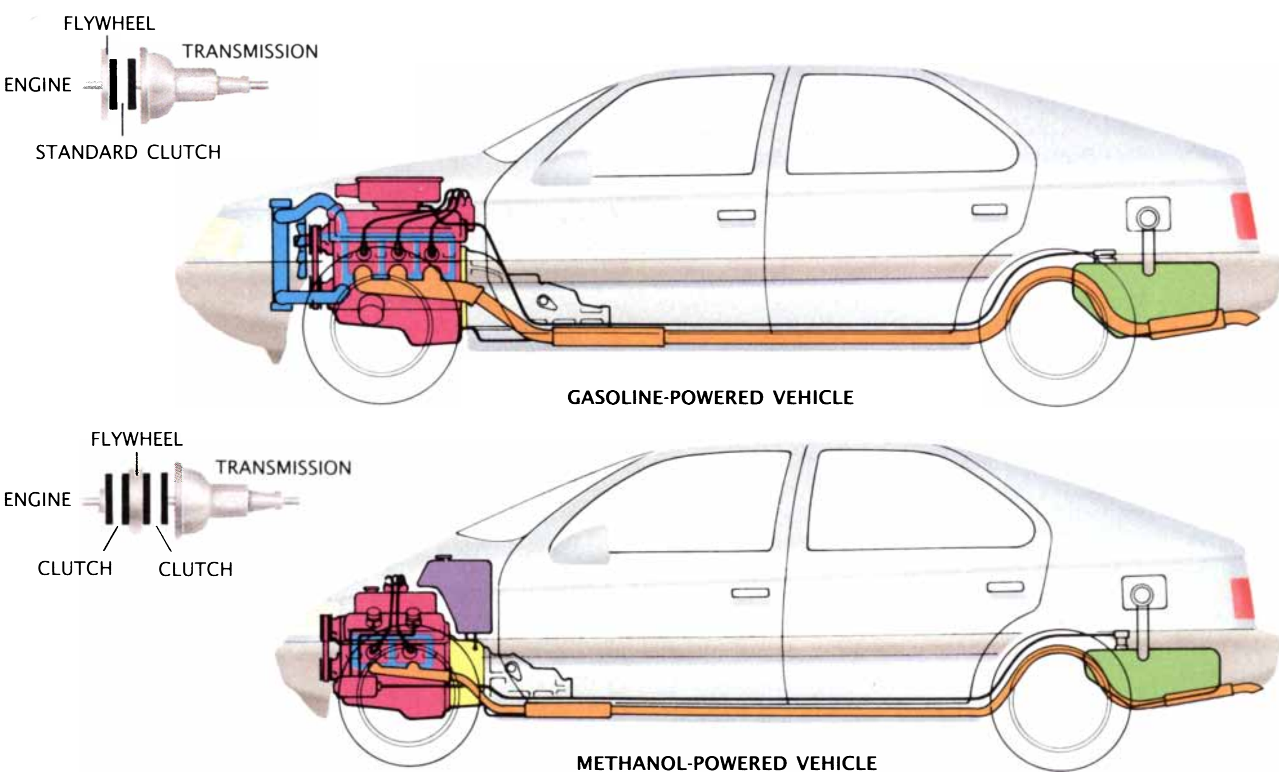
Like any alternative fuel, methanol has its critics. Yet many of the criticisms are based on "gasoline clone" vehicles that do not incorporate even the simplest design improvements that are possible with methanol. It is true, for example, that a given volume of methanol contains only about half the energy that gasoline and diesel fuel do; other things being equal, the fuel tank would have to be somewhat larger and heavier. But because methanol vehicles could be designed to be much more efficient, they would need comparatively less fuel. There are a variety of other objections, which we shall address later.

The first methanol-fueled production vehicles are likely to incorporate only the simplest of the engine improvements that methanol makes feasible. Even so, these vehicles would immediately lessen urban air pollution. Based on tests carried out with prototypes at the EPA, these early methanol-powered vehicles would also be about 30 percent more efficient than the best gasoline-engine technology and would achieve superior vehicle acceleration.

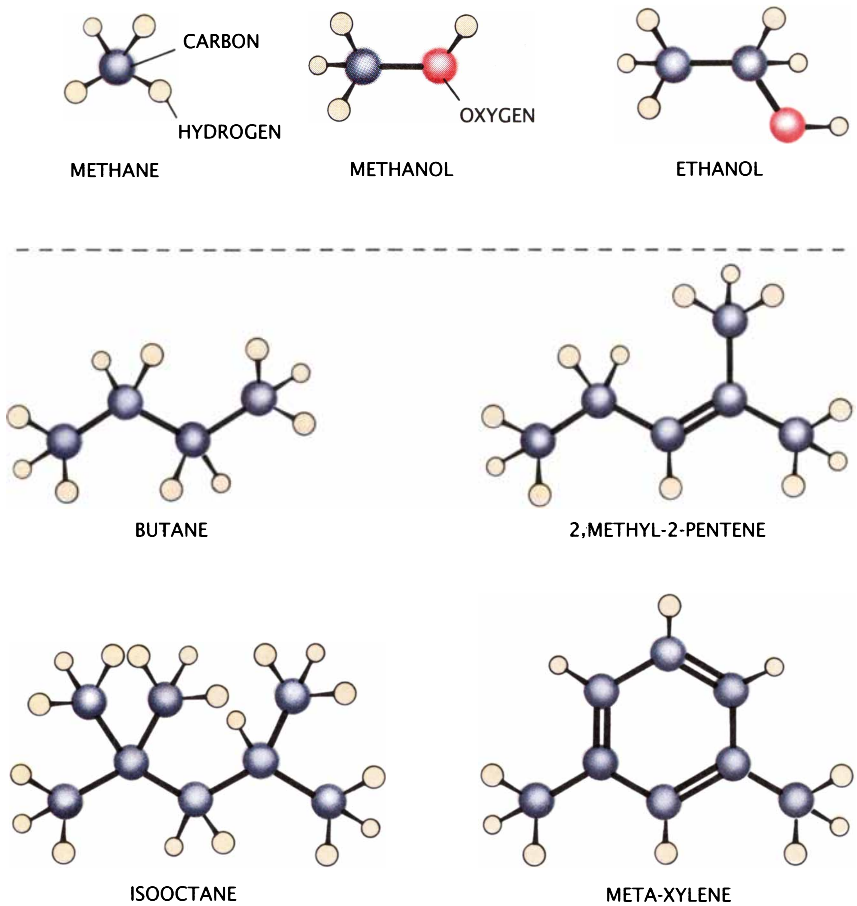
But the methanol automobile of the future would incorporate several additional features that cannot currently be attained with gasoline or diesel fuels. (Other features we describe could be employed in vehicles powered by gasoline or diesel but not while maintaining the low emissions of methanol.) These improvements would boost the overall efficiency of the vehicle, leading to even greater reductions in carbon dioxide emissions than would be attained simply by redesigning an engine to burn methanol. They would, of course, greatly enhance the mileage obtainable from a gallon of methanol.

To understand where these improvements will come from, one first needs to consider how conventional automotive engines work. Internal-combustion engines convert the chemical energy stored in fuel into mechanical energy: combustion produces hot gases, which exert pressure on the walls of the combustion chamber. The pressure forces a piston to move, turning the crankshaft and ultimately the drive wheels.

Several factors decrease the efficiency of the engine. Heat escapes through the walls of the combustion chamber, thereby reducing the pressure in the chamber and hence the energy available to do work. Furthermore, at the end of the expansion phase, which is limited by physical constraints on the engine design, the combustion gases are still at high temperature and pressure. The hot gases are exhausted from the engine even though they



METHANOL-FUELED CAR could integrate various features to attain higher efficiency and generate fewer emissions than a conventional gasoline-fueled car. The high octane and low heat loss of methanol combustion make possible an efficient and small engine (*red*), reducing the size of the fuel tank (*green*), exhaust pipe (*orange*) and transmission (*yellow*) as well. The cooling system (*blue*) could be made smaller by replacing the radiator and fan with thermal insulation and a hotter coolant. This change would decrease the size and aerodynamic drag of the front end. A flywheel start-stop system (*detail*) could shut down the engine whenever the car slowed down. A hydraulic pump-motor (*purple*) could store energy during braking.



SIMPLE HYDROCARBON FUELS such as methane, methanol and ethanol burn cleanly, forming mostly carbon dioxide and water. Gasoline molecules, on the other hand, are complex and contain many bonds between carbon atoms. Such molecules are more likely to leave uncombusted and photochemically active hydrocarbon compounds.

still contain a significant quantity of energy.

The engine of a passenger car, then, converts at most about 35 percent of the chemical energy in the fuel into useful work. In typical city driving, the engine averages only about 15 percent. On a combination of city streets and highways, about 30 percent of that energy goes to acceleration, 30 percent to overcoming rolling resistance and 30 percent to overcoming aerodynamic drag; the remaining 10 percent accounts for frictional losses in such moving parts as the transmission and for operating accessories.

What can be done to enhance the efficiency of a car? First, one must increase the fraction of the fuel's chemical energy that the engine transforms into mechanical work. Frictional losses in the moving parts, including the tires, must be minimized, as must the aerodynamic drag. Because both the frictional losses and the energy required for acceleration are directly proportional to vehicle weight, reducing that weight must be a high priority. Finally, one must reduce the energy that is lost when the brakes are applied, perhaps by storing the deceleration energy for later use.

A methanol-fueled car enjoys several advantages when it comes to boosting efficiency. To begin, one can design a car that would weigh significantly less than a conventional gasoline- or diesel-powered car having the same performance capabilities. Several features of the methanol engine make this weight reduction possible. The methanol engine, like the diesel engine, would inject fuel directly into the combustion chamber, rather than through a carburetor or air-intake manifold. Diesel fuel combusts spontaneously under compression, and so it must be injected at the top of the piston stroke against very high pressure—approaching 20,000 pounds per square inch—which requires an expensive and heavy injection system.

Methanol, in contrast, has a very high octane value (meaning it is more difficult to ignite through compression) and so can be injected under low pressure in the early phase of the compression stroke. The methanol engine can thereby make do with a simpler and less costly injection system. In addition, the injection can be timed to provide the most efficient combustion for varying engine loads. Later injection would create a stratified charge (a rich, easily ignited mix in one part of the chamber and a lean, cleaner burning mix in the rest) for low-power operation. Earlier injection, on the other hand, would give the fuel and air time to mix completely to achieve homogeneous combustion, the most efficient process for high-power operation. The properties of diesel fuel, in contrast, make homogeneous combustion impossible. What is more, because of the carbon-carbon bonds in the fossil-fuel molecules, the power output for a diesel engine (or for a conceptual direct-injection gasoline engine) is severely constrained by the need to keep smoke emissions within acceptable limits.

Because direct injection of methanol introduces the fuel at the spark or glow plug, it enables the engine to be started rapidly even in very cold climates, a problem presented by methanol engines that introduce the fuel through an air-intake manifold. As mentioned previously, methanol's high octane means that it will not ignite spontaneously even under very high compression. The engine can therefore be highly turbocharged or supercharged: the intake air can be compressed within the cylinder even before the compression stroke begins without causing the methanol to ignite. This reduces the work the engine must do to pull air into the cylinder. It also makes more air available for combustion. Indeed, methanol provides such excellent power output and efficiency that it is the only fuel used in Indianapolis-500 auto racing.

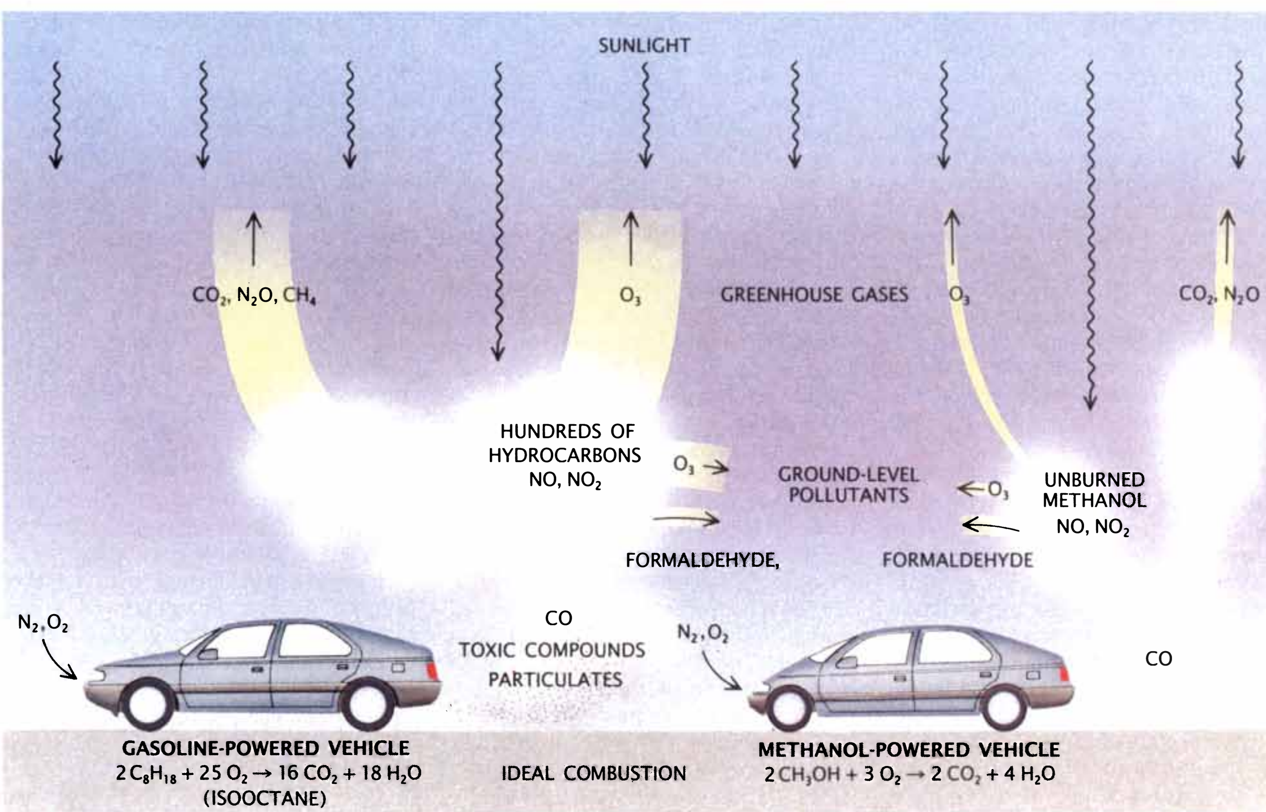
Another important feature of methanol is that it reduces heat loss during combustion. There are three principal reasons. First, methanol requires 10 times as much heat to vaporize as an equivalent quantity of gasoline; consequently, the peak temperature of the combustion gas is lower—and it is this temperature that drives the heat loss. Second, because methanol has no carbon-carbon bonds, it produces essentially no soot. Because carbon particles are very effective at radiating heat, their absence significantly reduces radiant-heat loss. Third, methanol has a higher ratio of combustion-gas molecules to reactant molecules than does gasoline or diesel. This improves the efficiency with which heat expands the gas and also lowers the temperature of the combustion gases, because there are more combustion-gas molecules to be heated by the energy released during combustion.

Methanol's low heat loss has the obvious advantages of improving efficiency and increasing the heat energy retained in the exhaust gases (which can be recovered for useful work by a method we shall discuss later). Less obviously, this feature can further reduce vehicle weight: because methanol reduces heat loss substantially, the engine's cooling system can be changed. Indeed, the radiator and cooling fan could be eliminated altogether if one applied thermal barriers, such as ceramics, to the combustion-chamber walls and reduced the temperature difference between the combustion gases and the engine coolant by employing hotter coolants (of around 300 to 400 degrees Fahrenheit). Many researchers have tried to apply this idea to gasoline and especially diesel engines, but no practical design has emerged because of the high combustion temperatures and large heat loss inherent to these fuels.

Eliminating the radiator would have important benefits in the methanol car: the car would weigh less, and overall it could be redesigned to reduce aerodynamic drag. (Conventional cars must compromise aerodynamic design to accommodate the radiator and a large surface area to force air through the radiator and engine compartment.) All told, then, the weight associated with the engine, including its cooling system and structural components, would be only half that of a diesel engine and two thirds that of a gasoline engine.

With a lighter engine and cooling system, it would be possible to lighten other parts of the methanol vehicle. A lighter engine means the vehicle could have a lighter frame, lighter suspension system, lighter brakes and lighter wheels. Indeed, for every one-pound reduction in engine weight, there would be a further three-quarter-pound reduction in the weight of other parts. A lighter vehicle would need less energy to accelerate or to overcome tire-rolling resistance and internal frictional losses, and so as the weight of the vehicle went down, the engine could be made still lighter.

The car we envision would combine a light, efficient methanol engine with a five-speed manual transmission or a continuously variable automatic transmission. The engine would be coupled to the transmission through a dual-clutch, computer-controlled flywheel. Apart from the vehicle's aerodynamic contour, the driver would not notice anything unusual about the car until the key is turned in the "ignition." The key activates a small electric motor, which engages the flywheel. Within five seconds, the flywheel is rotating at 2,500 revolutions per minute, after which the electric motor disengages, leaving the flywheel spinning freely. As long as the key remains in the "on" position, the electric motor comes on periodically to boost the flywheel back to 2,500 r.p.m. Until the driver decides to move the car, the engine is not run-



EMISSIONS from future methanol vehicles could contain one fifth as much carbon dioxide and one tenth as much of various hydrocarbons as emissions from today's gasoline-burning vehicles. Methanol vehicles would almost eliminate emissions of particulates and toxic compounds. Hydrocarbons and oxides of nitrogen react photochemically to make ground-level ozone, an irritant of the upper respiratory tract. Ozone, carbon dioxide, methane and nitrous oxide are greenhouse gases.

	METHANOL	ETHANOL	COMPRESSED NATURAL GAS	LIQUEFIED PETROLEUM GAS	ELECTRICITY
FEEDSTOCK SIZE/DIVERSITY	++	–	++	–	++
ENVIRONMENTAL IMPACTS	++	++	++	++	++
VEHICLE COST	0	0	–	–	– –
VEHICLE UTILITY (range, luggage space)	0	0	– –	0	– –
VEHICLE PERFORMANCE	0 / +	0 / +	– –	–	– –
CURRENT FUEL OPERATING COST (low demand)	–	– –	0	0	0 / +
FUTURE FUEL OPERATING COST (high demand)	++	–	+	0	0 / +
REFUELING CONVENIENCE (time, complexity)	0	0	– –	– –	– –

++ MUCH BETTER THAN GASOLINE
\+ SOMEWHAT BETTER THAN GASOLINE
0 SIMILAR TO GASOLINE
– SOMEWHAT WORSE THAN GASOLINE
– – MUCH WORSE THAN GASOLINE

ALTERNATIVE FUELS' environmental, economic and performance characteristics are compared with those of gasoline made from crude oil. It is assumed that vehicles would be designed to achieve optimal efficiency and performance with each fuel.

ning, and so consumes no fuel and generates no pollution.

To start the engine, the driver presses the accelerator pedal, which signals the computer to engage the engine to the flywheel. The engine begins to run almost immediately. (Direct injection of the fuel eliminates normal starting delays.) If the driver's foot lifts completely off the accelerator pedal, the computer disengages the flywheel, and the engine stops again. To move the car, the driver runs the engine, moves the gearshift lever to drive or first gear and further depresses the accelerator pedal. If the driver depresses the pedal slowly, the computer releases the clutch gradually, and the car accelerates gently. If the pedal is depressed rapidly, the computer sends more fuel to the engine, the clutch releases quickly and the car accelerates rapidly.

Apart from the sensation of going from a "dead" engine straight to acceleration, the car responds like a conventional automatic-transmission car. The version equipped with a continuously variable transmission behaves as though it had a normal automatic transmission, except that the engine always runs at its most efficient mode. For the manual-transmission version, the driver must shift gears. There is no clutch pedal, however, because the car has an automatic clutch. To achieve peak acceleration, the driver must fully depress the accelerator. This will engage a belt-driven supercharger, which will generate a palpable surge of power.

When power is temporarily not required of the engine—for example, when the driver observes a stoplight ahead—releasing the accelerator causes the computer to disengage the engine from the flywheel, and the engine stops running. This may be a little disconcerting at first, since the sound of the running engine and the feel of the engine slowing the vehicle will be absent; the vehicle will impart a sensation of "sailing." When the car slows to a flywheel speed of 2,500 r.p.m., the second clutch is activated, and the flywheel spins free of the decelerating transmission. As the car brakes to a stop and "idles," only the flywheel remains spinning, "storing" energy for starting the engine when the light turns green. Diesel and perhaps direct-injection gasoline engines could employ this flywheel system, but the engines and vehicles would become much heavier, and most important, certain emissions, such as particulates, oxides of nitrogen and ozone-forming hydrocarbons, would increase.

Compared with the average vehicle of today, the improvements discussed above could reduce the energy loss (and greenhouse emissions) for a four-passenger methanol car by between 60 and 80 percent. Much of the remaining loss results from braking and the ejection of hot exhaust gases. To make additional gains in efficiency and to reduce carbon dioxide emissions further, one must find ways to recover energy from these sources. One attractive scheme for recovering the energy dissipated by braking would be to install a variable-flow hydraulic motor-pump. The system would engage the flywheel during braking and pump fluid into a reservoir, thereby serving to brake the car and transfer kinetic energy from the car's motion into potential energy in the reservoir. Later the stored energy would be employed to spin the flywheel. If the pressure in the reservoir reached a sufficiently high level, the hydraulic energy could supplement the drive power of the engine. (A diesel- or gasoline-powered car could also employ such a system.)

Another avenue for improving efficiency lies in methanol's unique propensity to be catalytically dissociated at a relatively low temperature. As methanol flowed from the fuel tank to the engine, it would pass through a small dissociator, where heat from exhaust gases would be used to cleave the methanol into carbon monoxide and hydrogen, which actually contain 20 percent more energy than methanol itself. This reaction would take the exhaust heat and transform it into premium chemical energy. The enriched fuel would then be sent to the combustion chamber.

In considering the environmental impact of switching to methanol, one must also consider its manufacture. Natural gas would be an attractive feedstock: the gas that is now vented or burned off at remote oil wells could instead be converted to methanol, which is more economical to transport to market. Not only would such a scheme end the waste of this energy, but it also would eliminate a substantial source of greenhouse gases. In the short term, as long as petroleum prices remain low, methanol would probably be synthesized abroad, in the many countries abundant in natural gas. Unfortunately, that would mean continued dependence on imported energy. One interesting alternative would be to tap the large amount of natural gas produced on Alaska's North Slope; in fact, if Alaskan oil production slows down in the 1990's, the Alaska pipeline could transport methanol.

In the long term, we think that coal—of which there are vast reserves in the U.S.—could become a more attractive feedstock. This possibility has raised objections because the synthesis of methanol from coal produces higher levels of carbon dioxide than do other fossil-fuel feedstocks. There are several possible ways to reduce the carbon dioxide waste. One way is to employ more efficient processes, in which more of the carbon in coal is converted into methanol rather than carbon dioxide. A second way would be to collect carbon dioxide emissions at the plant, either for sale or for disposal in depleted oil and gas wells. Yet a third, long-term option would be to add hydrogen to the process: the excess carbon dioxide could then be turned into more methanol. The hydrogen could be supplied by electrolyzing water using electricity supplied by a nonfossil-fuel-power plant.

If coal-fed methanol plants were required to keep carbon dioxide emissions to a minimum, and if methanol vehicles were made to be highly efficient, it is projected that total carbon dioxide emissions might be as little as one fifth of the amount now generated by vehicles that burn gasoline refined from crude oil. If methanol were extracted from biomass, which absorbs as much carbon dioxide as it emits, no further carbon dioxide would enter the atmosphere, and contributions to global warming would be negligible.

The demand for methanol as an automotive fuel would also offer a new market for the high-sulfur coal that is now burned in electric-power plants in the Midwest and the East and that contributes greatly to acid rain. Instead of forcing plants to install costly scrubbers or putting thousands of coal miners out of work, policymakers could reduce acid-rain emissions by designating high-sulfur coal as a feedstock for methanol, whereas the low-sulfur coal reserves in the Southeast and West could fuel the nation's power plants. Such a program not only would supply the country with a cleaner transportation fuel and reduce acid rain but also would increase domestic employment and economic growth, reduce oil imports and foreign-trade deficits and enhance national security.

To put the cost of such a change into perspective, one might consider the following. If the high-sulfur coal burned in the U.S. were converted into methanol, that methanol could more than replace all the oil the U.S. imports from the Persian Gulf. And if the billions of dollars spent on the increased U.S. military presence in the Gulf could be spent on plants to convert coal into methanol, these plants could be fully funded in less than two years—and they would provide clean, domestic liquid fuel for decades.

We have highlighted many of the immediately foreseeable benefits of methanol fuel and have identified some of the areas for future technological improvement. With a change of such magnitude, however, the possibility of creating new problems must be thoroughly evaluated. We believe all the anticipated problems can be solved.

Critics note, for example, that methanol vehicles would emit twice as much formaldehyde as gasoline vehicles currently do. (The formaldehyde from vehicle emissions is estimated to account for about 75 cancer cases a year in the U.S.) Yet one must also consider that only one third of the formaldehyde related to vehicle emissions comes directly out of the tail pipe; the remaining two thirds is formed by the photochemical conversion of hydrocarbon emissions into formaldehyde. Pure-methanol vehicles will generate only one tenth as much of the hydrocarbons that are photochemically converted to formaldehyde as gasoline vehicles generate. The ambient formaldehyde level, then, could fall if vehicles were converted to methanol. In any case, the effect would be dwarfed by the impact methanol would have on other carcinogenic air pollutants, including diesel particulates, 1,3-butadiene, polycyclic organic compounds and benzene. Together these compounds are estimated by the EPA to account for about 800 cancer cases a year. If gasoline and diesel fuel were completely replaced by methanol, these toxic pollutants would be virtually eliminated.

Various other objections to methanol—that it is corrosive and acutely toxic, is explosive in a closed tank and has an almost colorless flame—are not considered serious obstacles by automotive engineers. To avoid corrosion, one would replace certain parts of the fuel system with materials such as stainless steel or inexpensive plastics such as polyethylene. To avoid the possibility of accidental ingestion, very low levels of colorants and odorants can be mixed with methanol (which is otherwise colorless and odorless). Baffles, or flame arrestors, would be placed in the opening of the fuel tank to prevent accidental ignition of methanol vapor. The flame arrestor would also prevent siphoning, which is the dominant cause of accidental ingestion of fuel. Incidentally, methanol vapor in the open air is far less flammable than gasoline, which makes it safer in spills and car crashes. In the rare vehicle fire, other substances besides methanol would be likely to burn as well, generating visi-

THE SYNTHESIS OF METHANOL

Methanol (CH_3OH) is synthesized by a catalyzed reaction of carbon monoxide with hydrogen.

$$CO + 2H_2 \xrightarrow{\text{CATALYST}} CH_3OH$$

The necessary carbon monoxide will most likely be provided by feedstocks of natural gas (methane) or coal. Natural gas (CH_4) reacts with water or oxygen to form carbon monoxide and hydrogen. Some carbon dioxide is also generated in the overall process that provides either the heat or the oxygen.

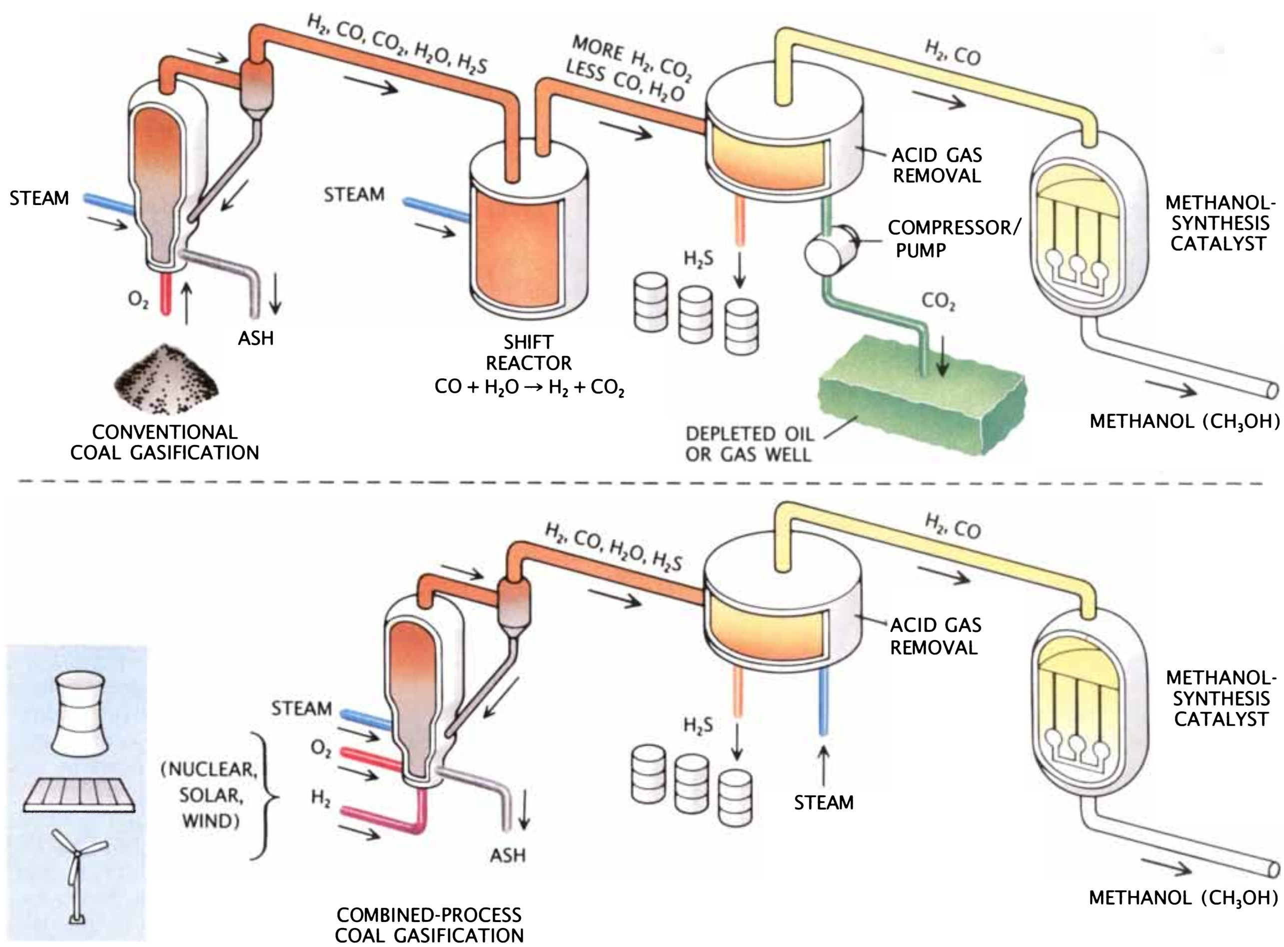
$$CH_4 + H_2O \xrightarrow{\text{HEAT}} CO + 3H_2$$

$$2CH_4 + O_2 \longrightarrow 2CO + 4H_2$$

To make methanol from coal, the coal must be gasified. The carbon (C) in coal is reacted with water and oxygen to form carbon monoxide, carbon dioxide and hydrogen. The process creates more waste carbon dioxide than does the methane-based process because coal has a higher ratio of carbon to hydrogen.

$$2C + H_2O + O_2 \longrightarrow CO + CO_2 + H_2$$

$$CO + H_2O \longrightarrow CO_2 + H_2$$



CARBON DIOXIDE is generated during the conventional synthesis of methanol from coal (*top*), both when coal is gasified to hydrogen and carbon monoxide and at a second stage, when steam is added to increase the ratio of hydrogen to carbon monoxide. Excess carbon dioxide, together with hydrogen sulfide produced from sulfur in the coal, must be removed before the mixture is passed over sensitive catalysts to yield methanol. The carbon dioxide could be pumped into a spent oil or gas well. Alternatively, the excess carbon (which would otherwise result in carbon dioxide) could be converted to methanol by adding hydrogen from an external source, such as the electrolysis of water at a nonfossil-fuel-power plant (*bottom*).

ble flames; but to allay concerns, one could employ a fuel additive that would make the flame visible.

There are compelling reasons for beginning the transition to methanol soon, because it will take many years for the program to reach a beneficial scale. Such a transition demands national leadership. The entire transportation infrastructure is set up for petroleum fuels, and the activities of major sectors of the economy will have to be coordinated if the transition is to succeed. Although it may be in the national interest to develop alternatives to petroleum fuels, it may not be in the immediate interest of individual companies or sectors that depend on the current infrastructure. Industry does not want to build cars for which fuel is not yet widely distributed; gas-station owners do not want to carry a fuel for which there is not yet a demand. They must be persuaded that methanol is a viable alternative fuel, and they will need incentives to make the necessary changes.

Although this is not the venue for an extended discussion about the economics and public-policy issues of switching to alternative fuels, we do want to make two important points. First, it should be remembered that there is a precedent for successfully introducing a new transportation fuel throughout the country: the federal government, automotive industry and oil industry worked together to introduce unleaded gasoline in 1975. Second, many European countries tax gasoline at levels that if they were applied here not only would encourage the development of alternative fuels but also would help to alleviate federal budget and trade deficits. It is time for the U.S. to begin designing a methanol-based transportation system for the next century. By doing so, the country would take a major step forward in solving its environmental problems and securing its energy destiny, and it would stake out a leading role in planning the future of global transportation.

FURTHER READING

MOVING AMERICA TO METHANOL. Charles L. Gray, Jr., and Jeffrey A. Alson. University of Michigan Press, 1985.

AUTOMOTIVE METHANOL VAPORS AND HUMAN HEALTH. Special Report of the Institute's Health Research Committee. Health Effects Institute, May, 1987.

NEW TRANSPORTATION FUELS. Daniel Sperling. University of California Press, 1988.

QUANTITATIVE ESTIMATE OF THE AIR QUALITY IMPACTS OF METHANOL FUEL USE. Armistead Russell et al. Prepared for the California Air Resources Board and the South Coast Air Quality Management District by Carnegie-Mellon University, May, 1989.

ELEK-TEK 10th year 1979–1989

Great Products at Great Prices from a Great Company

HUGE DISCOUNTS ON TOSHIBA EPSON OKIDATA HEWLETT PACKARD IBM / Diablo

Boards & Chips

Intel Aboveboard

Above Board Plus w/512K ... $399
Above Board Plus I/O w/512K ... 450
Above Board 2 Plus for PS/2 models 50, 50Z & 60 w/512K ... 475
Inboard 386 PC w/1MB ... 599
Inboard 386 PC 1MB Piggyback ... 345

AST Boards

SixPak Plus w/384K ... 279
RAMpage Plus 286 w/512K ... 399
RAMpage Plus 286 w/2.5MB ... 725
5251/11P Terminal Emulator ... 525

Chips and Simm Modules

256K 1 x 80NS ... 6.00
256K 1 x 100NS ... 4.50
256K 1 x 120NS ... 4.50
256K 1 x 150NS ... 4.50
Simms (all speeds) ... CALL

Math Coprocessors

Intel 8087 ... math (5MHz) ... $85
Intel 80872 ... math (6/8 MHz) ... 120
Intel 80871 ... math (8 MHz) ... 159
Intel 80287 ... math (6/8MHz) ... 125
Intel 802878 ... math (8 MHz) ... 189
Intel 8028710 .. math (10 MHz) ... 219
Intel 80C287 ... for 80C286 ... 265
Intel 8038716 .. math (16 MHz) ... 335
Intel 8038720 .. math (20 MHz) ... 375
Intel 8038725 .. math (25 MHz) ... 475
Intel 8038733 .. math (33MHz) ... 575

Video Cards

PARADISE

Paradise Short Color RGB ... $49
Paradise Mono Graphics ... 79
Paradise EGA 350 ... 109
Paradise EGA 480 ... 119
Paradise VGA (Basic) ... 149
Paradise VGA Plus ... 179
Paradise VGA Plus 16 ... 249
Paradise VGA Professional ... 299

Mono Hercules Compatible ... 49
Hercules Color ... 149
Hercules VGA ... 189
Hercules Graphics PLUS ... 179
Everex EGA 640 x 480 ... 129
Renaissance VGA I ... 169
Renaissance VGA II ... 219

Power Protection

Tripplite(TL) & Perma Power(PP)

Surge Protector Strips

TL Isobar 4 outlet ... $42
TL Isobar 6 outlet ... 49.99
TL Isobar 8 outlet ... 57
TL Isobar 8 outlet w/remote ... 65
PP PR610 6 outlet ... 16.99
PP RV610 6 outlet ... 28
PP RT410 4 outlet ... 29.99
PP RT610 6 outlet ... 31.99
PP PT209 2 outlet ... 19.99
PP PTD209 2 outlet with fax/modem protect ... 39

Command Consoles

PP RS562 5 outlet ... 65
TL CCI6-12 6 outlet ... 69
TL CCI6-plus with fax/modem protect. 75

Line Conditioners/Stabilizers

TL LS600 2 outlet, 600 watt ... 85
TL LC1200 4 outlet, 1200 watt ... 145
TL LC1800 6 outlet, 1800 watt ... 199

Standby Power Systems

TL BC325 watt system ... 269
TL BC450 watt system ... 299
TL BC1200 1200 watt system ... 650
TL BC2000 2000 watt system ... 1199
PP SPS500 400 watt system ... 499

ALL TRIPPLITE & PERMA POWER SUPPLIES IN STOCK AT DEEP DISCOUNT

Drives & Backups

Hard Drives

Seagate

20 Megabyte 65 MS ... 219
20 Megabyte w/Cont. XT ... 269
20 Megabyte 3.5" XT w/Cont. ... 309
30 Megabyte RLL XT w/Cont. ... 289
40 Megabyte 28 MS 1/2 HT ... 359
80 Megabyte 28 MS ... 599
123 Megabyte 28 MS w/Cont. ... 875

Miniscribe

30 MB RLL w/Cont. ... 289
40 MB 1/2 HT 28 MS ... 429
70 MB 1/2 HT 22 MS ... 599
150 MB ESDI 17 MS ... 1099
338 MB ESDI 16 MS ... CALL

WESTERN DIGITAL

Hard Drive Controllers

WD1003-MM1 MFM 2:1 HD only AT ... 99
WD1003-MM2 MFM 2:1 HD/Flop AT ... 109
WD1003-SR1 RLL 2:1 HD only AT ... 109
WD1006-MM1 MFM 1:1 HD only AT ... 119
WD1007-WAH ESDI HD only AT ... 150
WD1004A-WX1 MFM HD only XT ... 49

Hard Drive Cards

WD30 IFC 30 MB ... 299
WD40 IFC 40 MB ... 399

Hard Drive Cards by Plus Development

20 MB Hard Card ... 520
40 MB Hard Card ... 645

Tape Backups

Mountain Internal

40 MB XT or AT ... 379
60 MB AT ... 899

Toshiba Disk Drives

ND 04D 1/2 HT 360K (PC/XT) ... 79
ND 04DEG 1/2HT 360K (PC/AT) ... 79
ND 08DEG 1.2MB (PC/AT) ... 89
ND 352KU 3.5" 720K (XT/AT) ... 89
ND 356T 1.44MB (PC/AT) ... 115

Monitors

Zenith

14" VGA Flat Screen ... $599

Samsung

Flat-screen Amber ... 115

Packard Bell

12" Amber ... 79
14" Amber Flat Screen ... 99
14" White Flat Screen ... 105
14" CGA Color ... 229
14" EGA Color ... 335
14" Uniscan Color ... 375
14" VGA Color ... 385

NEC

Multisync 2A color ... 499
Multisync 3D color ... 649

Amdek

210 Plus Mono *New* ... 109
410A, 410G, 410W 12" Mono ... 150
632 14" Color VGA ... 375
732 12" Color VGA ... 399
735 14" Smartscan Color VGA ... 499

Sony

CPD 1304 14" Multiscan Color ... 660

3M Computer Tape

2400' polyback-w/tape seal ... $10.78 ea.
2400' Wrightline-w/tape seal ... 11.85 ea.
2400' Blackwatch-self-threading ... 12.30 ea.
3600' Blackwatch-w/tape seal ... 17.75 ea.
3600' Blackwatch-self-threading ... 18.18 ea.

(Sold in cases of 10 only)

Printers

EPSON LX-810 $189

- Affordable and professional quality
- Proven 9-pin technology

Epson Printers

FX850/1050 ... CALL
DFX5000 ... CALL
LQ510 *NEW* ... CALL
LQ850/950/1050 ... CALL
LQ2550 ... CALL

Toshiba Printers

P312 SL/321 SLC ... CALL
P341 SL ... CALL
P351 SX ... CALL
ExpressWriter 301/311 ... CALL

Okidata Printers

172/Parallel *NEW* ... 199
182 Turbo Parallel *NEW* ... 240
320 ... 350
321 ... 499
380 *NEW* ... 375
390 ... 499
391 ... 699
393 ... 999
393 (color) ... 1099

Accessories

Print Buffers

Practical Peripherals

64K parallel or serial ... 145
256K parallel or serial ... CALL
64K parallel interface for Epson ... 75

SMG

PB36 256 parallel ... 199

Rotary Switch Boxes

488024 2 pos. 25-pin AB ... 15
488058 2 pos. 36-pin AB ... 18
448028 3 pos. 25-pin ABC ... 25
448060 3 pos. 36-pin ABC ... 28
448078 4 pos. 36-pin ABCD ... 30
Other Switch Boxes ... CALL

HEWLETT PACKARD Genuine DeskJet/LaserJet Supplies

Trays

Letter Size for LaserJet II/IID ... 52
Legal Size for LaserJet II/IID ... 60
Envelope Size for LaserJet II ... 70
Letter Size for LaserJet IIP ... CALL
Legal Size for LaserJet IIP ... CALL
Envelope Size for LaserJet IIP ... CALL

Laser Toner Cartridges

HP 92275A LaserJet IIP* ... CALL
HP 92285A LaserJet* ... 84
HP 92295A LaserJet II/IID* ... 84

DeskJet Ink Cartridges

DeskJet/Plus Ink Cartridges* ... 14.99

Memory Boards

1MB for LaserJet IIP ... CALL
2MB for LaserJEt IIP ... CALL

*Quantity Discounts Available

TRONICS 2400 Baud Modem

• 100% Hayes Compatible • 2 Yr. Wty.

Internal w/ Procomm software List $199 **$99**
External no software List $239 **$129**

1200 Baud Internal w/software CALL

Mouse Devices

Numonics Cordless Mouse ... $99
Logitech Hi-Res Bus or Serial ... 89.99
Logitech Hi-Res PS/2 Mouse ... 74.99
Keytronics Professional Mouse ... 69.99

Microsoft

Paintbrush Bus or Serial ... 98
Paintbrush w/Windows Bus or Serial ... 135

Mouse System

White Mouse w/paint Bus/Ser ... 69.99
Omnimouse II w/paint Bus/Ser. ... 49.99
PC Trackball w/magician Bus/Ser. 64.99

Scanners

Logitech Scanman ... $225
DFI Handscanner HS-3000 Plus ... 199

Modems

Hayes and Hayes Compatible

Hayes Internal (including software)

1200 Baud ... $279
2400 Baud ... 429

Compatibles (including software)

Practical Peripherals 1200 Baud ... 59.99
Practical Peripherals 2400 Baud ... 129
Practical Peripherals 2400 MNP/5 ... 175
Everex 2400 Baud ... 129

Hayes External (no software incl.)

1200 Baud ... 277
2400 Baud ... 416

External Compatibles

Practical Peripherals 2400 Baud ... 179
Practical Peripherals 2400SA MNP/5 199
U.S. Robotics Courier 1200 ... 179
U.S. Robotics Courier 2400 ... 299
U.S. Robotics Courier 2400E ... 385
U.S. Robotics Courier HST ... 599

For Toshiba Portables by PR & D

2400 Baud Internal Modems

For T1000 ... 175
For 1100+, 1200, 1600, 3100, 5100. 189

Fax Machines

AT&T

3410D/3520D ... $899/1299
4525D ... CALL

Murata

M900 ... CALL

Ricoh

RF800/850 ... 679/750
RF 900 ... 850
RF 920 ... 1175

Toshiba

Model 3300 ... 750
Model 3600/3700 ... 899/1049

Sharp

UX180 ... 750
UX350 ... 1250
FO220 ... CALL
FO330 ... CALL

We stock all popular FAX PAPER at deep discount prices

3M Data Cartridges

Regular		Pre-formatted	
DC100A(Mini)	14.50	DC600 HC/Iotamat™	25.50
DC300A	17.90	DC615 HC/Iotamat™	19.00
DC300XL/P	21.50	DC6150/Zeta™	28.00
DC600A	23.00	DC1000/Alpha™	15.90
DC600HW	23.00	DC1000/Delta™	15.90
DC6037	18.00	DC2000/Gamma™	21.50
DC615A	17.50	DC2000/Kappa™	21.50
DC6150	24.00	DC/2000/Theta™	21.50
DC 6320	29.00	DC2080/Rhomat™	22.50
DC1000(Mini)	14.50	DC2110/Numat	24.90
DC2000(Mini)	17.00		
300/600 Series	19.99		
Data Cart. Clean. Kit			

Call for Large Quantity Pricing

Diskettes	Dysan	ELEK-TEK	IBM	maxell	MEMOREX	SONY	3M	Verbatim
3 1/2" SSDD	-	-	-	11.50	10.99	11.00	11.50	-
3 1/2" DSDD	11.50	10.99	18.00	12.00	10.99	12.00	13.00	9.99
3 1/2" DSDD preformatted	-	-	-	-	-	-	13.00	-
3 1/2" DSHD preformatted	-	-	-	-	-	-	31.00	24.99
3 1/2" DSHD	26.00	23.99	34.00	27.00	27.99	27.00	31.00	23.99
5 1/4" DSDD	6.99	5.99	10.00	6.99	4.99	-	7.50	4.99
5 1/4" DSDD preformatted	9.00	-	-	-	-	-	7.50	8.50*
5 1/4" DSHD preformatted	14.00	-	-	-	-	-	14.00	14.00*
5 1/4" DSHD (IBM AT)	12.99	9.99	23.00	14.00	9.99	-	14.00	12.00

* Teflon Coated

ELEK-TEK, inc.®

7350 North Linder Avenue, Skokie, Illinois 60077

Call Toll Free **800-621-1269**
Prices subject to change
all products subject to availability.

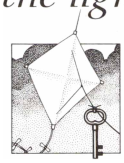
Call in IL **312-677-7660**
Canadian Toll Free **800-458-9133**
We Offer Large Quantity Bid Opportunities

Corporate accounts invited, open accounts shipped F.O.B., Skokie, IL, prepaid orders, minimum order $15.00 Visa or Mastercard by Mail or Phone. Mail cashiers check, money order, personal check, add $4.00 1st Item. (AK, HI, P.R., Canada add $10.00 1st item) $1.00 ea. add'l. shpg. and handl. Shipments to IL add 7% tax. **Write for free catalog.** IBM is a trademark of International Business Machines Co. Epson is a registered trademark of Seiko Epson Corporation.

ALL ELEK-TEK MERCHANDISE IS BRAND NEW, FIRST QUALITY AND COMPLETE.
Return Policy: All sales are final except defectives which will be replaced with identical merchandise only. Computers and large peripherals – 3 work days after delivery, most small items, 30 calendar days, after which manufacturer's warranty applies.

THE AMATEUR SCIENTIST

Colored segments of a grid can shed a diffuse glow like the light from a neon tube



by Jearl Walker

"Neon spreading" is not a warning that gas is escaping from a Times Square signboard. Rather, the term was introduced in 1975 by Harrie F.J.M. van Tuijl of the University of Nijmegen in the Netherlands to describe a certain illusory spreading of color resembling the diffuse light that issues from a glowing neon tube.

The human visual system is remarkably skilled at reducing the bewildering amount of information received by the eyes to a recognizable mental picture of the external world. Yet in some circumstances the picture contains curious errors: visual illusions, which often offer a glimpse into otherwise hidden operations of the visual system. In some illusions the failure of the visual system to create a precise mental picture of a rather spare diagram can be quite perplexing.

One of van Tuijl's demonstrations of the neon-spreading illusion is a case in point: the illustration is simple, and yet we perceive it erroneously [*see illustration at left, below*]. To see the illusion, hold the page upright, with the horizontal lines level, and avoid fixing your gaze on any part of the illustration. (It may also help if you hold the page at somewhat more than the usual reading distance.) Bright blue-tinted spots will appear to overlie the blue crosses, as though some of the blue ink from the crosses had seeped out over the surface of the page. The spots may also seem to connect to form slanted lanes. Why do the colored spots appear?

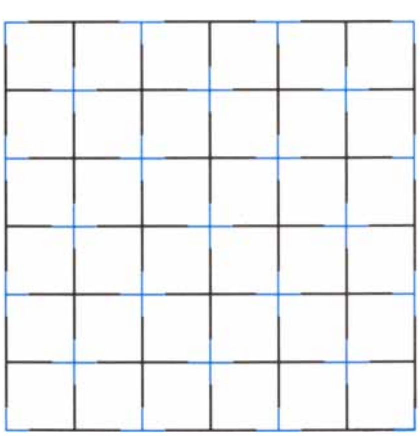
The neon-spreading illusion is related to another illusory appearance, called subjective contours, which was the subject of this department in January, 1988. You can produce an example of subjective contours by redrawing the grid with blank spaces where there are now blue crosses. In the blank spaces you will perceive bright circular spots with perceptible borders. The bright spots are not really there, of course; they are said to be subjective. In van Tuijl's illustration, the blue of the crosses spreads out and fills the circles. The illusory bright spots in the empty gaps and the neon color surrounding the colored crosses may be of similar origin, because they appear in the same kind of pattern.

Van Tuijl offered two other patterns, reproduced at the right, below, to demonstrate that the neon spreading is not ubiquitous in grid patterns. In the figure on the left the red bleeds into the framework of the design, but in the one on the right the red is confined to the lines. Interconnections, present on the left but not on the right, must be somehow important.

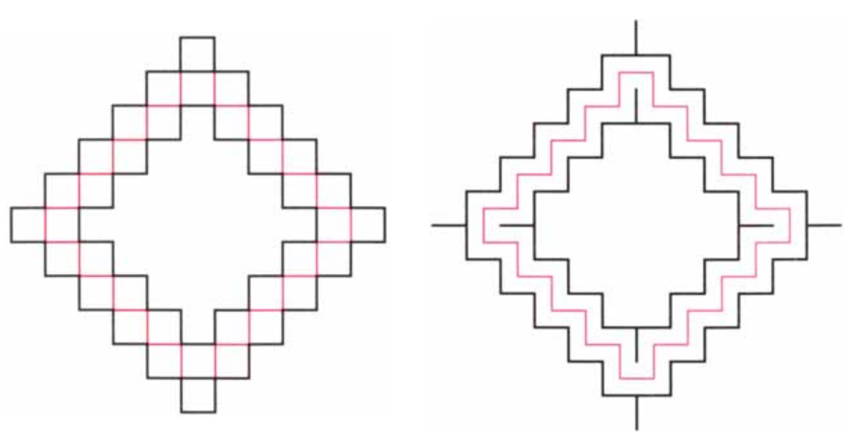
In 1979 van Tuijl and his colleague Charles M. M. de Weert examined additional grids, including some in which bright lines lay on a dark background. Three rules emerged from their work. If neon spreading is to appear, both the basic grid and the interrupting crosses must be either lighter or darker than the background. If they are both darker than the background, as in the illustration at the left, below, then the crosses must be less dark than the grid (blue rather than black, in this case). If they are both light on a dark background, the grid must be lighter than the crosses. In either case, then, the brightness of the crosses must be intermediate between the brightness of the grid and that of the background.

Although the rules allow one to predict whether a particular grid pattern will display neon spreading, they do not reveal the source of the illusion. According to van Tuijl and de Weert, the illusion stems from the way the visual system interprets what is being seen: the system apparently finds it more "efficient" to perceive faint, colored spots and dulled, murky crosses instead of the true scene. The argument was repeated in another 1979 paper by van Tuijl and E.L.J. Leeuwenberg, also a colleague.

The concept of efficiency here may need clarification. As an example, consider again that first illustration below. When you view it, information about the black lines and colored crosses is



Van Tuijl's grid



***A grid that yields the illusion* (left) *and one that does not* (right)**

sent along the pathway from the eyes to the visual cortex of the brain. Somewhere along the route the system begins to try to make sense of what is being seen. It compares the relative brightness of background, grid and crosses, and it also considers the overall pattern, in particular its repetitive elements. The system might liken the pattern to what has been seen in the past. It might consider whether parts of the pattern may be hidden by something closer to the viewer. After these and other matters are weighed, an interpretation is made and a final picture is brought to the level of consciousness.

When you look at the first of van Tuijl's illustrations, your visual system presumably decides that the most probable interpretation of it is that parts of what is actually a complete grid of black lines are masked or obscured by projections of circular blue spots. Where the projections seem to fall on the grid, the blue glow dulls the contrast between the colored crosses and the surrounding white region by tinting the white.

Why should the visual system be attracted to such an interpretation? Perhaps it is because the system seeks to complete line segments that extend toward each other the way the black lines do from each side of a colored cross. That would relate the pattern to other grids or gridlike structures often seen before—a fence, say. It is normal that an object that is closer to the viewer than a gridlike structure would be brighter than the grid; the fact that the crosses are brighter than the black lines supports the notion that they are in front of the grid and block the view. A cross is not commonly seen precisely positioned to obscure segments of a grid, however, and so the visual system conjures up a spot, which is a more likely obstacle. If the crosses are darker than the grid, the whole interpretation falls flat because the crosses do not appear to be in front of the grid; the attempt at mentally completing the grid fails, and there is no need for obscuring spots or neon spreading.

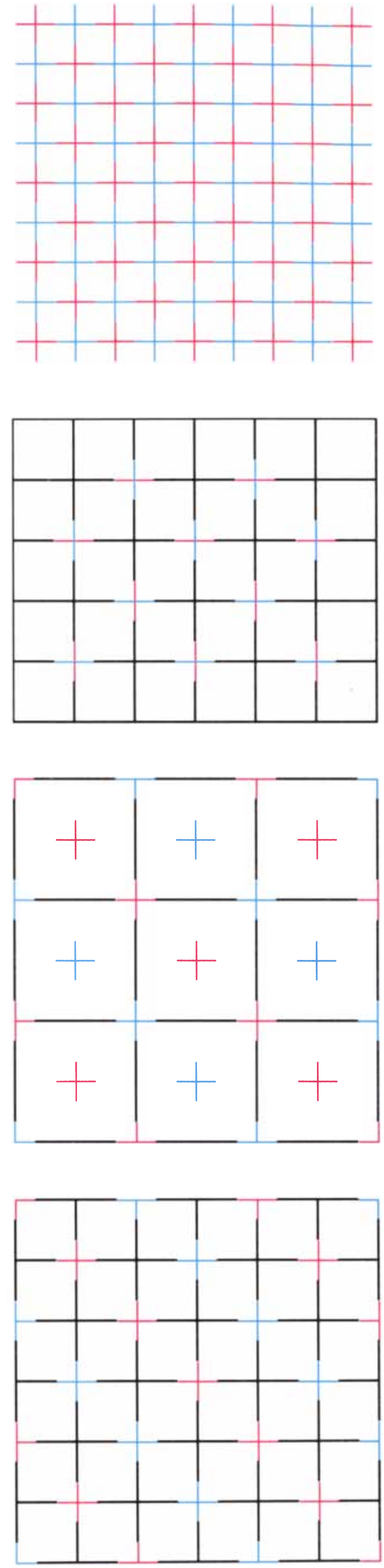
That explanation seems to me to be incomplete. I do not think I fancy a completed grid more than I do one with color changes, which is actually more interesting. I am also bothered by the fact that once I realize the illusion is an error, I cannot make the error go away. Even when I examine the illusion for a while, concentrating on the error, I still perceive the spread of color. If my visual system strives for efficiency and also compares a present view with learned information, surely after enough experience and conscious thought it should at least reduce, if not eliminate, the error of neon spreading. Could there be, instead, some stage in my visual system where a physical process introduces an error—one that is so well entrenched by the time it reaches the conscious level that neither visual experience nor willpower can shake it? If that is the case, where exactly does the erroneous processing take place?

A bountiful harvest of clues came in 1981, when Christoph Redies and Lothar Spillmann, then both of the University of Freiburg in West Germany, reported their studies of variations of the grid arrangements that give rise to neon spreading [*see illustration at right*]. In the top figure red crosses replace some grid crossings, and blue crosses replace others. Red neon spreading adorns the red crosses, blue neon spreading decorates the blue ones and red or blue slanted lanes connect the colored spots. To see the illusion best, you should again hold the page so that the horizontal lines of the grid are level and adjust the distance between your eyes and the page. If you then rotate the page around your line of sight, the illusion weakens as the rotation reaches 45 degrees and reappears as the previously vertical grid lines become horizontal.

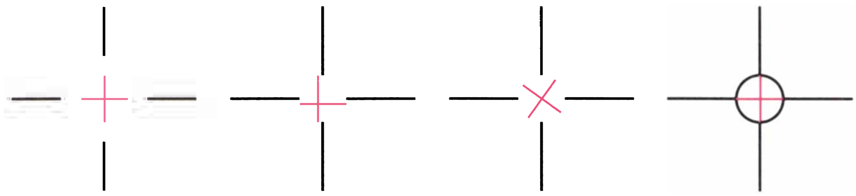
Now move back from the grid until the blue crosses grow dark enough to blend with the black lines they connect. Although you can no longer distinguish the color of the crosses, they still induce what are perceived as being blue spots. If the crosses were yellow instead of blue, moving back would cause them to fade into the background until they disappeared, but yellow spots would still be recognizable. Either experiment demonstrates that although the color of the crosses may not survive to the conscious level, it can continue to trigger neon spreading.

When red or blue crosses are isolated from the grid lines, as in the second figure from the top, they fail to produce neon spreading at any angle of orientation; the juxtaposition of the grid's black lines with a cross must, then, be vital to the illusory spread of color. Notice also that an isolated cross is more sharply defined and contrasts more with the background than does one enmeshed in the grid: as in previous examples, neon spreading makes the cross look dull and murky.

Two more variations are seen at the right. In the third figure from the top each cross is made up of one red line



Variations on illusory grids



Variations on a grid unit

Variations on lines and crosses

and one blue line. Colors spread from both colored lines but not as much as in the previous examples. In the bottom figure the black lines are eliminated, so that the grid consists solely of red or blue crosses. Yet the neon spreading of each color is robust and even yields slanted colored lanes. The result adds a new rule about brightness. In this case either the red or the blue crosses can be considered the "grid," and the crosses of the other color are taken to be the "crosses." Neon spreading can apparently appear when both elements are about equal in brightness but are darker than the background.

Redies and Spillmann investigated how the strength of neon spreading depends on the angular size of the colored crosses enmeshed in a black grating. When the pattern is on the direct line of sight (which is called foveal viewing), the angle occupied by a cross in an observer's field of view has to be between four and 35 minutes of arc if neon spreading is to appear. At the lower end of the range, the spreading is strongest and fills out a circle. At the upper end, it is weaker and takes on a diamond shape. When the angle is even larger, the spread of color retreats and lies just along the sides of the lines in the cross, an effect dubbed neon flanks. When the pattern is off the direct line of sight by several degrees (extrafoveal viewing), the upper and lower ends of the angular range for neon spreading shift to somewhat larger values. All of this indicates why the illusions on these pages often intensify if you adjust your distance from the illustrations: you thereby vary the angle subtended in your field of view by each colored line or cross, optimizing the illusion.

Redies and Spillmann also examined variations of a single unit in which one cross has black lines at each end [*see top illustration on this page*]. When the cross is disconnected from the black lines by a separation or by misalignment through displacement or rotation, neon spreading disappears. The spreading is also eliminated if the cross is circled. If the black lines are short, the illusion is strong; it may still appear, albeit weakly, even if there are only mere dots at the ends of the cross.

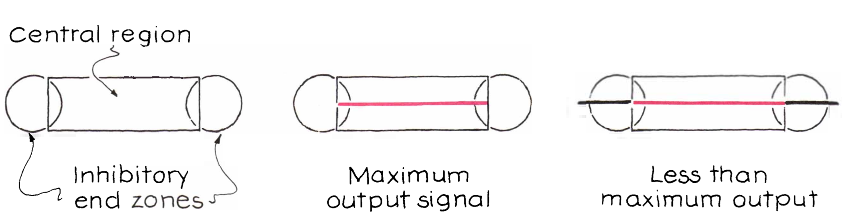
The most intriguing illustrations by Redies and Spillmann are the ones in which they strip the illusion to the bare essentials: single lines [*see middle illustration on this page*]. If an isolated red line is viewed, there is no spread of its color, but if it connects two black lines so as to form a single, straight line, red neon flanks run along the red line. The spreading of color disappears if the black line on either side is removed.

With the results from these and other experiments, Redies and Spillmann conjectured that neon spreading is a result of the activity of visual cells that detect the presence and orientation of lines and that are sensitive to line lengths corresponding to the angular ranges given above. Apparently the neon spreading takes place only if the line segment has straight continuations at each end and if the continuations differ from the segment in brightness or color. If the continuations are misaligned, the averaging and spreading disappear; they also disappear if there is only one continuation. When a colored cross replaces an intersection in a black grid, the color spreads from both lines of the cross, but the spreading is more extensive than just the sum of the neon flanks along the lines. This result suggests that the visual cells detecting each arm of the cross may interact.

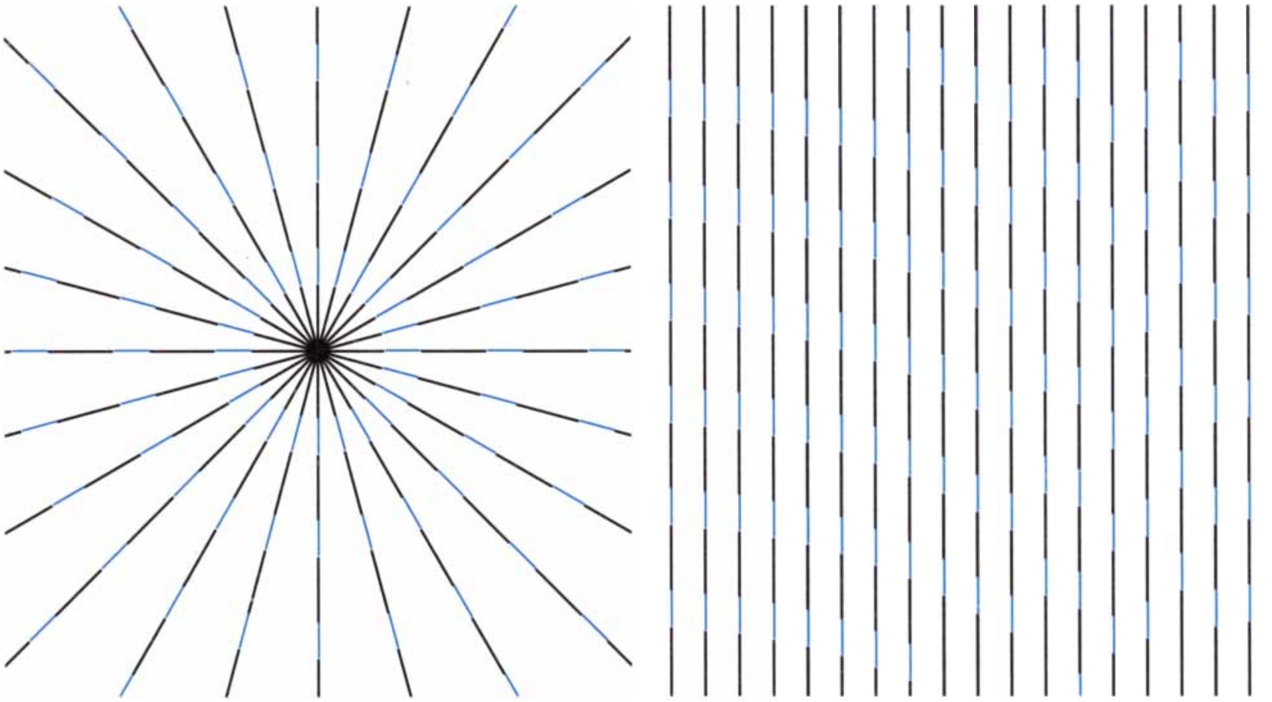
In 1984 Redies, Spillmann and their associate Kristian Kunz continued the study of neon flanks. Two of their figures appear on the opposite page. In the one at the left, notice that near the center of the star the blue lines are close enough for the color to spread between them, making a blue ring; farther out the blue lines provide only neon flanks. Neon spreading can also be seen in the figure at the right wherever the blue segments either are lined up or are shifted from one another in an orderly way. Where the positions of the segments are scrambled, you see only neon flanks.

Following clues from other studies of vision, Redies, Spillmann and Kunz proposed that the "end-stopped" cells in the visual cortex are responsible for the dulling of a colored line when neon spreading or neon flanks appear, and that they may figure in the entire illusion. Most cells in the visual cortex have the assignment of detecting lines, but the end-stopped cells have the particular assignment of detecting short lines. The visual field of an end-stopped cell has an elongated central region with inhibitory zones at each end [*see bottom illustration on this page*]. If both end zones are activated, they inhibit the activity of the central region. The design renders the cells sensitive to short lines, making them promising candidates in explanations of neon spreading.

When a line excites a cell, the line is said to have been "projected" onto it. (To be sure, the only true optical image lies on the retina; only electrical signals reach the brain. But the idea of



Responses of an end-stopped cell



Patterns that yield both neon spreading and neon flanks

projection simplifies the description of how a cell works.) Suppose that a short red line on a white background lies on your direct line of sight and that its projection is on one of the end-stopped cells that monitor foveal viewing. If the projection is skewed with respect to the long axis of the cell, the cell ignores it, but if the projection is aligned with the axis, the cell sends a signal deeper into the visual system. When the projection occupies only the central region of the cell, the signal is maximum, conveying the information that a dark line lies on a brighter background. What you then perceive is a short red line that contrasts well with the background.

Now suppose that the red line has black continuations at each end and that the continuations pass through the end zones of the cell. The end zones inhibit the discharge triggered by activation of the central region, and so the output signal is less than maximum, which indicates that the line is not as dark as it was when it stood alone. What you now perceive is a red line that has less contrast with the background: it is dulled, as is the case when either neon spreading or neon flanks appear in the illusory illustrations.

An end-stopped cell may dull a line, but can it also spread the color perpendicularly to the line? No one knows yet, but a few months ago Redies wondered whether an end-stopped cell might interact with other, adjacent cells to spread the color of the line onto the adjacent background. One check will be to determine how large an angle a line can occupy before its projection exceeds the central region of an end-stopped cell. Does the angle match the upper limit of about 35 minutes that Redies and Spillmann reported for neon spreading in foveal viewing?

Other questions also remain. When a black grid with colored crosses is tilted, why does neon spreading weaken? Does the result reveal that end-stopped cells respond primarily to horizontal and vertical lines? Why does the spreading also disappear if a cross is encircled with a black line? And what visual cells are responsible for subjective contours?

It seems to me that the brightness rules for neon spreading can be explained by the end-stopped-cell model, if one assumes that the extent of the spreading is related to the amount of inhibition by the end zones. For example, recall that for the illusion to appear on a bright background, the colored lines must not be as dark as the black lines. To see why, again consider a short red line with black continuations. The fact that the continuations are quite dark means that the end zones diminish the output from the cell to a great extent, which may mean that the neon spreading and the dulling of the red line are enough to be perceptible. Now consider how a short black line with red continuations affects a cell. The central line, being quite dark, tends to make the cell fire strongly, and the continuations, being lighter than the central line, produce only a small inhibition—and so also less neon spreading and less dulling of the central line. Although this line of reasoning is interesting, it has not yet been experimentally tested by neurophysiologists.

FURTHER READING

A NEW VISUAL ILLUSION: NEONLIKE COLOR SPREADING AND COMPLEMENTARY COLOR INDUCTION BETWEEN SUBJECTIVE CONTOURS. H.F.J.M. van Tuijl in *Acta Psychologica*, Vol. 39, pages 441-445; 1975.

THE NEON COLOR EFFECT IN THE EHRENSTEIN ILLUSION. Christoph Redies and Lothar Spillmann in *Perception*, Vol. 10, pages 667-681; 1981.

COLORED NEON FLANKS AND LINE GAP ENHANCEMENT. C. Redies, L. Spillmann and K. Kunz in *Vision Research*, Vol. 24, No. 10, pages 1301-1309; 1984.

CORTICAL DYNAMICS OF THREE-DIMENSIONAL FORM, COLOR, AND BRIGHTNESS PERCEPTION: I. MONOCULAR THEORY. Stephen Grossberg in *Perception & Psychophysics*, Vol. 41, No. 2, pages 87-116; 1987.

DISCONTINUITIES ALONG LINES: PSYCHOPHYSICS AND NEUROPHYSIOLOGY. Christoph Redies in *Neuroscience and Biobehavioral Reviews*, Vol. 13, No. 1, pages 17-22; Spring, 1989.

COMPUTER RECREATIONS

A microgolf game gives professionals and amateurs an equal chance for a hole in one

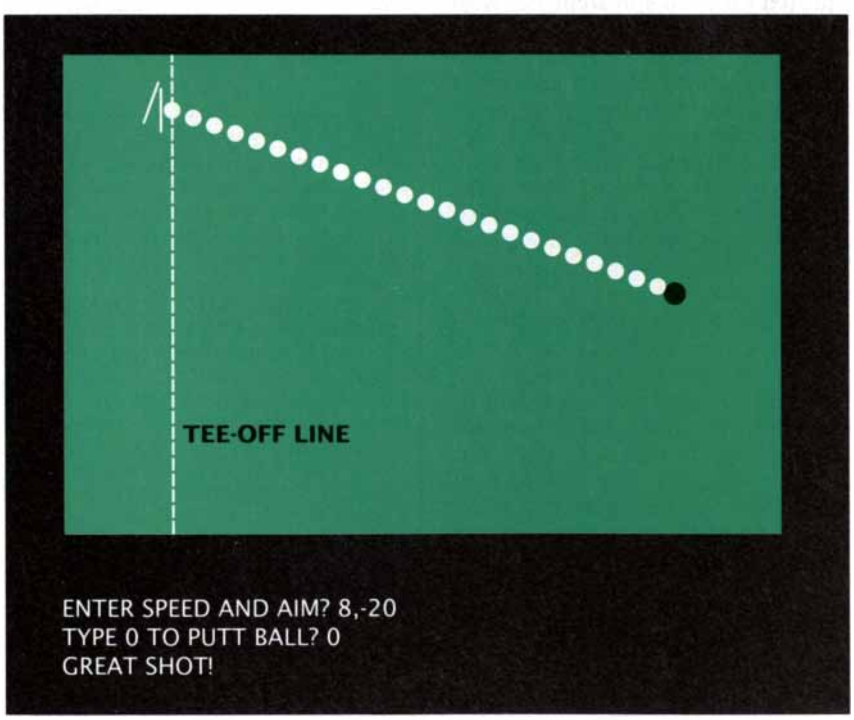
by A. K. Dewdney

"I have once, it is true, had the distinction 'of making a hole in one'.... That is to say, after I had hit, a ball was found in the can, and my ball was not found. It is what we call circumstantial evidence—the same thing that people are hanged for."

—STEPHEN LEACOCK,
"Mathematics for Golfers," in *The World of Mathematics,* Volume 4

The ordinary golfer deploying an armory of clubs tries desperately to drive a small ball into a distant hole hidden in several acres of grass and guarded by ponds, trees and sand traps. Players of miniature golf use but one club on a much smaller obstacle course of bumps, tunnels, rotating blades and swinging pendulums. Now players of microgolf, without swinging a club at all, can attempt a hole in one on the screen of a microcomputer. There are hazards here, too, including some that violate the laws of physics.

Although the microgolfer is confined to an armchair, he or she can putt with an electronic club and even enjoy a dimension not available to players of the larger games: programming. How many of us are able to design, build and play a course all in the same day? Not even Jack Nicklaus—unless he knows how to program. In fact Nicklaus would not even have to be a professional programmer to set up a microgolf course.



A successful putt on the HOLE IN ONE green

I shall lay down plans for three holes: one for beginners and two for more advanced players. At the end of the day everyone who plays by the rules will have made a microgolf course on a par with his or her talents.

The project is illustrated in its simplest form on this page. Here the program I call HOLE IN ONE has displayed a single hole on a rectangular green. A putter appears as a short line segment behind a tiny ball at one end of the green. A player angles and positions the putter to aim and strike the ball toward its target, the cup at the other end of the green.

The simple HOLE IN ONE version of microgolf is all hit or miss. If the ball goes past the cup, it will cross over the course's edge as though it were not there, disappearing off the screen and rolling into the computer's memory, never to reappear. Even this version of the game has a certain pleasurable tension to it.

It is actually possible for an amateur to prepare HOLE IN ONE as a kind of software springboard to the more advanced versions. Fragments of the program HOLE IN ONE can be inserted into the programs I call BIRDIE and EAGLE, which are depicted on pages 121 and 122, respectively.

BIRDIE features a hazard near the cup, a circular twilight zone of sorts. If the ball enters this zone, it changes direction and speed in a completely unpredictable manner. More treacherous hazards plague the EAGLE green, but I will withhold the horrifying details for now.

Every year or so I have endeavored to provide a more detailed program description for those hesitant or beginning programmers who need just a bit of extra information to get started. The following description of HOLE IN ONE adopts the tried-and-true method of starting with a clear description of the computation to be performed, usually given in steps as an algorithm. From there it moves, as fast as possible, to actual fragments of program. A reader who puts all the pieces together is just a few keystrokes away from the *Scientific American* Open!

HOLE IN ONE first displays the cup ready for play, and then it requests that a player adjust the putter and putt. HOLE IN ONE then draws and redraws the ball as it rolls across the electronic turf, perhaps into the hole.

From these specifications a crude algorithm for HOLE IN ONE can be devised: draw the layout, prompt the user to putt, move the ball many times (to animate the direction and speed of the putt) and decide each time whether the ball has fallen into the cup. The algorithm can then be refined to create the program HOLE IN ONE. To begin with, the initial layout can be displayed in a few steps that draw the green, cup, ball and putter. These steps can be programmed easily.

In order to be helpfully explicit, I shall assume that the reader is writing a program in the BASIC language on an IBM PC. (Do not be discouraged if you lack the hardware. The program is easily adapted to other computer systems.) To be even more explicit, I shall pretend that all readers have a screen at least 300 pixels (screen dots) wide and 200 pixels high. All distances on the screen must be measured from the origin, which is the top left corner of the IBM screen.

On a 300-by-200 pixel screen it is quite reasonable to draw a rectangular green 240 pixels long and 160 pixels high. To center the layout more or less on the screen, HOLE IN ONE places the green 30 pixels from the left side of the screen and 20 pixels from the top. In short, the horizontal coordinates of the green run from 30 to 270 and its vertical coordinates run from 20 to 180. The green takes shape from the following instructions:

```
10 CLS
20 LINE (30,20) – (270,20)
30 LINE (30,180) – (270,180)
40 LINE (30,20) – (30,180)
50 LINE (270,20) – (270,180)
```

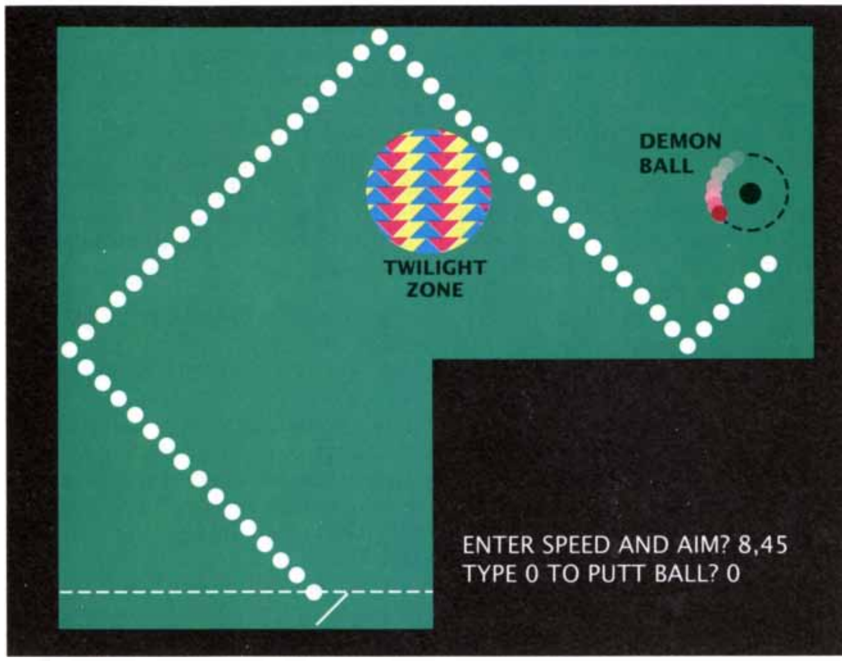
Note that by numbering BASIC statements with increasing (but not necessarily consecutive) values, the programmer designates the exact order in which the computer should execute those statements. The first command, CLS, clears the screen of any images that may previously have been drawn on it. This feature is employed when one wishes to restart the program.

HOLE IN ONE next represents the cup as a circle at the far right end of the rectangle. The center of the cup is 240 pixels from the left and 100 pixels from the top. The BASIC instruction that will draw the cup is therefore

```
60 CIRCLE (240,100),5,1
```

The circle is centered on the coordinates (240,100) and has a radius of five pixels. The "1" at the end of the line specifies the color white.

Just to keep the shot from being boring, HOLE IN ONE places the ball at a random location along the tee-off line, which is 30 pixels from the left end of the green. Because the ball's position changes throughout the program, the



A twilight zone and a demon ball make life difficult for the intermediate putter

variables *XBALL* and *YBALL* are created to keep track of its coordinates.

```
70 XBALL = 60
80 YBALL = 160 * RND + 20
90 CIRCLE (XBALL,YBALL),4,1
```

Line 70 restricts the ball's *x* coordinate to coincide with the tee-off line. Line 80 selects a random number between zero and one called RND, scales it up to the size of a number between zero and 160 and then adds 20 to the result. This means that *YBALL* will lie somewhere along the tee-off line. Line 90 draws the ball with its center at (*XBALL*,*YBALL*). The ball has a radius of four, a snug fit in the radius-five hole.

A line just behind the ball represents the putter, or more accurately the blade of the putter. Its resting state is given by the following:

```
100 LINE (54,YBALL + 8) –
    (54,YBALL – 8)
```

This vertical line is tangent to the left side of the ball and is 16 pixels long.

The next step of the HOLE IN ONE algorithm—to "prompt the user to putt"—must be refined just a little before programming continues. "To prompt" means not only to send a message that asks a player to putt but also to acquire information about the orientation of the putter and the speed of the swing. It also means that HOLE IN ONE must redraw the putter in its new position so that a player can judge the shot by eye.

Two lines of BASIC are added to HOLE IN ONE, both to print a message and to accept input.

```
110 PRINT "ENTER SPEED AND AIM"
120 INPUT SPEED, AIM
```

The variable called *SPEED* stores the distance in pixels that the ball moves every time HOLE IN ONE updates its position in the course of a putt.

The variable *AIM* stores the ball's direction of motion. Many players will find degrees to be the natural unit for entering an angle for *AIM*. HOLE IN ONE therefore accepts an angle between +90 and −90 degrees as measured from a horizontal line. The extremes represent strokes that send the ball straight up or down.

Unfortunately, most versions of BASIC handle angular measurements in units called radians rather than in degrees. Therefore, HOLE IN ONE requires a small calculation that converts degrees into radians.

```
130 RADAIM = (AIM * 3.1416)/180
```

The conversion is based on the fact that 180 degrees equals π radians.

The values of *RADAIM* and *SPEED* can be inserted into a formula that determines the position and orientation of the putter. HOLE IN ONE requires the formula to draw the putter poised either to stroke the ball into the cup or to whack it off the green.

The formula positions the center of the putter behind the ball along the stroke line and at a distance proportional to the value of *SPEED*. The putter should just touch the ball's surface when *SPEED* equals zero, that is, it should be at least three pixels to the left of the ball's center. The *x* coordinate of the putter's center is therefore discovered by subtracting (*SPEED* + 3) times the cosine of *RADAIM* from *XBALL*. The *y* coordinate will be found by adding (*SPEED* + 3) times the sine of *RADAIM* to *YBALL*.

To angle the putter, HOLE IN ONE requires a little more trigonometry. Since the putter is 16 pixels long, its end points are eight pixels away from the center. The displacement along the *x* axis is then eight times the cosine of *RADAIM*, and along the *y* axis it is eight times the sine of *RADAIM*.

These quantities, denoted by *A*, *B*, *C* and *D*, respectively, in the program fragment below, are all computed separately on lines 140 to 170. The actual coordinates of the putter's "top" end (*XTOP*,*YTOP*) and "bottom" end (*XBOT*,*YBOT*) are then computed. Finally, in line 220 the putter is drawn as a line connecting these points.

```
140 A = 8 * COS(RADAIM)
150 B = 8 * SIN(RADAIM)
160 C = (SPEED + 3)COS(RADAIM)
170 D = (SPEED + 3)SIN(RADAIM)
180 XTOP = XBALL - C + B
190 YTOP = YBALL + D + A
200 XBOT = XBALL - C - B
210 YBOT = YBALL + D - A
220 LINE (XTOP,YTOP) - (XBOT,YBOT)
```

It might appear at this point that the program is ready to enter the third phase of its operation, to produce the animation of the ball heading toward the cup. Has anything been left out? What if a player decides that the angle of the putter looks wrong. Surely the fallible human being should be given a second chance. This is done by branching back to line 10 at the player's option.

```
230 PRINT "TYPE 0 TO PUTT BALL"
240 INPUT PUTT
250 IF PUTT <> 0 THEN GOTO 10
```

Here, if the player types any number but zero, the program will branch back to line 10, where it will clear the screen, redraw the green and prompt the player again for new values of *AIM* and *SPEED*. If the player types zero, HOLE IN ONE will proceed to the final phase of its operation, the one specified much earlier by the clause "move the ball many times to animate the direction and speed of the putt."

To move the ball, however, HOLE IN ONE must increase the *x* and *y* coordinates of the ball independently. To this end, the variable *SPEED* is decomposed into two new variables, one called *XSPEED* in the *x* direction and one called *YSPEED* in the *y* direction.

```
260 XSPEED = SPEED * COS(RADAIM)
270 YSPEED = SPEED * SIN(RADAIM)
```

The final section of the program consists of a loop within which two operations will be performed continually. The ball will be moved and then checked to see whether it happens to be in the cup.

```
280 FOR K = 1 TO 300
290 XBALL = XBALL + XSPEED
300 YBALL = YBALL + YSPEED
310 CIRCLE(XBALL,YBALL),4,1
320 IF ABS(XBALL - 240) > 4
    THEN GOTO 360
330 IF ABS(YBALL - 100) > 4
    THEN GOTO 360
340 XSPEED = 0
350 YSPEED = 0
360 NEXT K
```

The instruction on line 280 sets up the simplest kind of loop. A variable *K* counts from one to 300 to ensure enough move-and-draw cycles for the ball to make it to the cup even at a speed of one. Within the loop the new ball coordinates are updated, and at line 310 the ball is drawn.

Lines 320 and 330 test the ball's coordinates separately to find out whether the ball is in the cup. If it is not, the program skips down to line 360. If it is, the ball's two speed coordinates are reset to zero, effectively freezing the ball.

Professionals may object to this "for" loop because the program continues to run even when the ball drops into the cup. HOLE IN ONE tests the ball's coordinates until the count *K* reaches 300. To remedy this, the amateur golf programmer might add three instructions to the end of the program. One would test whether the speed of the ball is zero. If the test is affirmative, a second instruction could print a pat-on-the-back message such as "GREAT SHOT!" A final instruction, required by most versions of BASIC, marks the end of the program with the word "END."

This completes HOLE IN ONE. Readers who type in the program and run it will notice a peculiar effect as they try to putt the ball at different *SPEED*'s. The unusual discrete physics that appears to operate in the microgolf world of HOLE IN ONE will produce a hole in one only if the speed of the ball divides the distance to the cup evenly!

What might be considered a defect from one point of view may be regarded as a "feature" from another; if the physics is wild by accident, perhaps it can also be so by design. BIRDIE features some hazards not seen on any grass turf. In addition, the ball bounces around the green.

To describe these golfing wonders, I will shift expositional gears to exclude IBM PC's and BASIC, reverting to algorithmic language. The BIRDIE course shown on the preceding page is L-shaped. How does BIRDIE ensure that

Three hazards on the advanced green

the ball stays within the borders as it rolls? Even in the strange world of microgolf, the angle of reflection must equal the angle of incidence. This is not difficult to set up for the six segments, which may be called *wall1* through *wall6* without worrying too much about which wall is which. Each time the main display loop calculates a new position for the ball, BIRDIE will check to see whether any of the walls has been crossed. For example, if *wall1* has been crossed, BIRDIE should reflect the ball.

When the test and its (possibly) resulting reflection have been executed, the program checks *wall2* and then *wall3* and so on. An interesting problem surfaces at this point. What if the ball, in "striking" one wall and then receiving its subsequent reflection, now finds itself outside another wall? Is there not a possibility that if the ball is destined to hit a wall near a corner, it will bounce only from one wall but not from the other? Readers who ponder the point properly will develop a successful solution.

Once the ball has done all the rebounding it is destined to do in the current cycle, BIRDIE displays it and tests whether the ball is in the cup just as HOLE IN ONE does. What if the unlucky putter misses the cup? The ball will continue to bounce around within the confines of the green until it either enters the cup or falls into an endless cycle of rebounds. Some duffers will find this a wonderful proposition; others will sneer at the lack of realism. The game lacks that all-important factor, friction!

To allow the ball to be slowed to a stop, BIRDIE multiplies *SPEED* by some constant that is less than one, say .995, each time the program goes through its move-and-test loop. Because *SPEED* decreases by a factor of .995 each time through the loop, both *XSPEED* and *YSPEED* will also be multiplied by this constant within the loop.

It may of course happen that the ball eventually slows to zero without having dropped into the cup. In this case BIRDIE must call a halt to the loop, and so the loop must be not of the "for" type but of the "while" type. The second kind of loop keeps the cycle going as long as (while) some condition holds. In this case the condition is that *SPEED* be greater than some rather small number such as .05. At such a point BIRDIE sends the user back to the interactive part of the program where he or she is prompted for a putt.

What hazards does the hapless golfer encounter on the second hole? One hazard is referred to as a circular twilight zone. Having entered the zone, the ball suffers a random change in its current direction. What happens is that BIRDIE tests whether the ball lies within the charmed circle and, if it does, changes the angle *RADAIM* by a random number between −10 and +10. By planning a careful series of bounces, a good putter can send the ball around the twilight zone.

The second hazard is harder to avoid. A demon ball orbits the cup at a fixed distance. It completes an orbit for every 10 cycles of the loop. If the player's ball happens to touch the demon ball, the player's ball reappears back at the tee-off line. Most readers who have followed the course this far will probably think of a way to manage this awe-inspiring event.

EAGLE, as the elaborate illustration on the opposite page shows, has a more complicated green than BIRDIE. Largely a glorified version of BIRDIE, it features some advanced hazards. For example, whenever the ball traverses the neck connecting one end of the layout to the other, it experiences a force pulling toward the neck's center. The force is related to the ball's distance from the centerline, as though the passage were gracefully banked. There is also a sand trap, where the *SPEED* is decreased not by a factor of .995 but by .9. Only the most careful of strokes will get one out of the trap without sending the ball careening around the layout like a runaway bullet. The final hazard involves a bit of "rough." Here the ball becomes lost, disappearing before it even rolls to a stop. How to find the ball? If you reposition the putter in the right location, the ball will appear beside it.

I was inspired to take readers on a tour of microgolf greenery by way of a game called Zany Golf, recently released by Electronic Arts in San Mateo, Calif. One can buy the game, of course, but that might mean missing the fun of building one's own. Readers are by no means limited to golf, in any case. The techniques described here lend themselves readily to games of micropinball and electronic billiards.

Connoisseurs of the Mandelbrot set, fans of fractals and a great many other readers eagerly devoured the July banquet of biomorphs, snails and fractal popcorn. The fisherman and chef who inspired the repast was Clifford A. Pickover of IBM's Thomas J. Watson Research Center in Yorktown Heights, N.Y. He creates a biomorph with a program based on essentially the same equation that generates Julia sets [see "Computer Recreations," SCIENTIFIC AMERICAN, November, 1987]. The insertion of extra conditions into the iterative process produces images that resemble microscopic views of living organisms.

David J. Hoffman of Greenville, Tex., notes that biomorphs, which contain higher powers of z in their formulas than Julia sets do, are generated more quickly. He sees biomorphs as Julia sets with "a little bit more" added to them. The extra pixels include not only the spikes and cilia but the cell body itself: "The central portion of the biomorph is either a connected Julia set or the fractal dust of an exploded Julia set."

Several readers encountered some trouble in generating the 12-pointed radiolarian that was illustrated on page 110. They obtained not the whole creature but a close-up of the cell body, so to speak. The window given was too small. Readers are advised to expand it from three by three to 20 by 20.

Strange that the plate should excite as much admiration as the food! Truchet tiles, our fanciful serving platters, turned out to interest almost as many readers as the biomorphs did. These tiles are squares decorated with two circular arcs that cut two opposite corners. An array of 0's and 1's can be converted into a crazy quilt of meandering curves by replacing each 0 with a tile in one orientation and each 1 with a tile in the other orientation. Pickover notes that even slight biases in the randomness of the bits can result in easily discernible patterns, a fact with implications for those who search data for patterns.

Joseph V. Saverino of Trexlertown, Pa., is a chemical engineer with an aesthetic bent. By varying the design on the tile, Saverino obtains a variety of results that range from strange urban-looking plans to random Moorish and Byzantine patterns. By flooding some of the connected areas with color, Saverino can instantly pick out the connected areas in his patterns.

Daniel C. Spencer of Valencia, Calif., has proposed an animated film program: once the random bits have all been made into tiles, switch randomly located bits one at a time and change the corresponding tile. Such a program, according to Spencer, should make "the entire screen . . . squirm and undulate," a mesmerizing effect!

FURTHER READING

MICROCOMPUTER DISPLAYS, GRAPHICS, AND ANIMATION. Bruce A. Artwick. Prentice Hall, 1984.

BOOKS

Usage, wetlands and dry, a long-ago SDI, the works of Babbage and the new physics

by Philip Morrison

DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY, by Sidney I. Landau. Cambridge University Press, 1989 (paperbound, $13.95).

This candid account of dictionary making not only takes us inside an old and important literary specialty but shakes the doubtful assumptions implicit in most everyday thought and speech. It is a surprisingly agreeable lesson in humility for all of us who enjoy those unexamined gifts of childhood, fluency in our mother tongue and correct grammar we don't even know. Sidney Landau is a savvy professional who learned how to define in the 1960's, working for Funk & Wagnalls in a big former stable on East 24th Street in New York City. Dictionary work, he says, is writing for a living to tight purpose, "not to be confused with *belles lettres*." But he is a genuinely philosophical thinker as well, apt at distinctions, aware of depth of meaning and of social nuance, always seeking evidence. His work opens a way into the science of language through some familiar and concrete applications; his book is a reflective handbook of linguistic engineering.

Language is alive in our daily speech, and so it is the dominant usages of speech that modern lexicographers have appealed to as standard for more than a century. But in fact what they know is almost entirely words in print, often from expensive and inadequate samples. For example, there is a widely used "Brown Corpus," prepared in Providence (and reviewed here) 20 years back. Passages—500 of them, of 2,000 words each—were taken from contemporary writings for adult readers over a wide range of topic and formality; 50,000 different words are listed, half of them appearing only once or twice. Words as familiar as *invert* or *ballistic* occur once only. A mere megaword corpus, then, can hardly be a guide to a large technical vocabulary. Yet 40 percent of ordinary dictionary words are technical.

It is not hard to compile words by the 10 million from one publication like the *New York Times*, but what real usage would so single a source represent, over the varieties of purpose, style and context of language? "I am skeptical that any ... class of people habitually use the vocabularies of the publications they habitually read.... We simply do not know what the habitual usages of most people who are not writers are," Landau writes, and dictionaries ought to admit it.

It is in documented quotations that lexicography watches the shift and flow of use and meaning, and a citation sample is the best source for new terms and senses. One knowing estimate reckoned that some 800 words enter the "common or working vocabulary" of English each year as others fade from currency into obsolescence. An old photograph in the book shows a studious old party, rather like white-bearded Father Time, peering at a paper as he stands in a room whose walls are lined by shelves of folded sheets; this was James A. H. Murray, chief editor of the unmatched *Oxford English Dictionary,* amidst some of the five million citations assembled long ago by two generations of well-educated Victorian amateurs. Their zeal built the basic dictionary in our language, the only one fully based on citations, however weakly it attended to any specialized vocabulary. That deficiency is structural, consistent with the generalist orientation of upper-class Britain, where it ruled the lexicographic tradition that was "not really breached" until the 1970's.

Dictionaries center on words, not on things. There are a dozen key elements, from the alphabetized order of main entries to the auxiliary matter front and back. Spelling is elementary and fixes entry order; English spelling was not firm before Samuel Johnson's time. It is a reviewer's bet that the schoolroom skill of spelling will more and more be passed off to cheap silicon memories. But as active spelling ability fades from readers' skills, its passive use to seek entries may seem more burdensome, and dictionaries will go on screen offering choices among alternative forms, perhaps also joining encyclopedic material that goes beyond artfully terse definitions and uses. Hints of the change are at hand, although it is true that "a book is a marvelous device for random access to short entries."

Definition is the heart of a dictionary. Here is a careful account of the logical issues. First of all comes circularity, as real a concern as it has been a logical quibble among schoolchildren. Then there is the rule against definitions that depend on words nowhere else defined. The "Word Not In rule" is broken more often than the circularity rule, for it is difficult to make sure that every word incorporated in any definition is itself defined somewhere. Nor is it right to define a concept rather than a word; lexicographic amateurs, often specialists brought in to muster the terms of one or another art, tend to confuse the two. The question at hand is: What does the word itself mean? Your definition must say something crisp about the word itself; you need to say that a diagnosis is a judgment about the nature of a disease, however neatly you describe how the physician reaches judgment.

Simplicity is one aim of artful definition. We dutifully laugh at Mr. Johnson's pompous definition of *network*: "... reticulated or decussated, ... with interstices between." But a good recent dictionary defined *feather* in five lines, not omitting a feathery hierarchy of structure, from barbs, barbules and barbicels down to tiny hamuli! Readers are likely to agree that the "meanings of many words, including ... deceptively simple ones, are seldom learned from dictionaries." But it is the interplay between lexical and ostensive or contextual meaning that tells us how each word functions, its implications and probable limits. Precision and simplicity are in conflict, but it cannot be forgotten that approximation and ambiguity are real parts of words' meaning. Here we touch the subtlety of that grand and flexible code that is natural language.

Taboo words are delightfully treated as problems, not without due attention to the serious issues they imply. The author cites a hilarious letter he once wrote on the word *f**k* (beginning with his plea that he could find no record of the actual four-letter word inquired after). "'Queen Isabella is said to have uttered it upon learning that the *Santa Maria* had foundered.'"

The rise of British lexicography to a new dominance during the 1980's is related. Its success is based on a wide and innovative attention to English as an international tongue, with books that provide original tools for second-language learners: grammar, pronunciation by the use of more elaborate phonetic script and, most freshly, a guide to word combinations—or idiom. They offer help enough not only to allow understanding but even to enable use. One important new family of books couches all its definitions within a limited vocabulary of 2,000 listed words; there is risk of dullness (all too well achieved in books for children that adapt such a technique), but there can also be a gain in understandability that may repay the motivated adult user.

Like good dictionaries, this book brings a reader too much to sum up. A critical bibliography of word reference books, extending well beyond ordinary monolingual dictionaries, is included. There are as well a brief history of the art, a fascinating discourse on the organization of an effective dictionary-making staff and a look at the future of computerization. A brief preface (the first edition appeared five years ago) brings us the author's current opinions.

MONSOON, by Steve McCurry. Thames and Hudson, 1988 ($35). **ETERNAL LANDSCAPE: UTAH, ARIZONA, COLORADO, NEW MEXICO,** by Emil Schulthess. Text by Sigmund Widmer. Alfred A. Knopf, 1988 ($60).

"For months there is no rain, then there is too much. Half the world's people survive at the whim of the monsoon winds. This is the reality, and it will stay with me forever.... In the heart of the monsoons, I was forced to immerse myself in weather so profound that nothing else mattered." So McCurry, a New York photographer of worldwide renown, recalls the half year when he followed the "festive disaster," the summer seasonal rains, across all of India, south to Sri Lanka and north to Nepal and Bangladesh, and overseas to the winter counterpart from Indonesia to northern Australia. The images he caught—wide landscapes of spring-green fields and the faces and postures of men, women and children both enraptured and beset by pouring water—give a memorable visual accounting of what too few outsiders see, here in the world where fair weather means famine.

Three descriptions suffice to outline

this work of infectious empathy. One strange photograph shows a bland watery surface from which only a disembodied head and a sewing machine emerge. Indeed, it is a grizzled man walking in water to his chin, carrying high a worn but quite dry sewing machine; he is tailor in a town in Gujarat in western India. His quiet, relieved and deeply warming smile discloses how pleased he is to have saved the means of livelihood from monsoon peril, which for him is plainly "not ... a natural disaster, but an annual event ... with a whiff of gaiety."

A wide landscape in Bangladesh stretches across another page—terrain flat, flat, flat to the edge of the view. Down the center runs a single track, draftsman-straight, whereon a red-tipped locomotive draws a dozen cars toward the airborne camera. Brimming waters spill over paths and roads and dikes, to stain and blur half of the neat rectangular boundaries of broad, green paddy fields, all in flood.

In a third picture a young woman in a bandbox-neat pink dress leans quietly and politely toward us across the top of the white picket fence at her front gate, her house a dozen feet beyond her. But the green surface that smoothly surrounds her image and fills the garden area is not grass beneath her feet; it is an unbroken layer of duckweed, and it is floating at waist height. For a fortnight the monsoon completely flooded her Javanese community. "The people were ... gracious.... They simply had nowhere else to go.... In Java, it was simply the season to be flooded out."

Each of these books provides its own intrinsic unity of image. The large pictures of the second book are as red and wide and unpeopled as the first book is at a glance green and intimate. Emil Schulthess is a celebrated Swiss landscape photographer who has traveled to the southwestern U.S. and recorded its extraordinary landscapes to complement his Alps. Nearly all the views are seen under wide skies at great distance, 10 of them in panoramic vistas carried large across a long gatefold of three pages, often from helicopter shots made with a remote-controlled rotating camera.

A few images must convey the flavor of this artist's deep look at much recorded scenes. Monument Valley is viewed in one double spread centered on the high red-mitten buttes we all recognize from 100 films. But the light is dimmer and more diffuse than we recall, and the monuments rise unexpectedly and mysteriously out of layered veils and swirling pools of white cloud. He had come to the land in a rare time of wet weather, the temperate monsoon in Arizona; a closer shot shows a torrent of red water cresting over a natural wall of desert land.

Across another gatefold, under a cloud-striped blue sky stretch the blue-white gypsum dunes of the White Sands. On another page we catch sight of Rainbow Bridge, largest of natural bridges, seen from high above amidst bare mountains; neatly framed by a dip in the stony rise we stand on, the big stone arch spans a narrow blue arm of Lake Powell. A shot of the tower of Spider Rock against the dizzying space of the Canyon de Chelly, where it casts its black shadow on crimson land, is the single photograph that most evokes the black-and-white images of the classical photographers of this celebrated land.

Both books provide helpful texts, introductory to what lies behind the images. These western lands are described in a simple physical way; a concise geological synopsis makes clear both the age of visible rock layers and the very different age of the landforms by which they are exposed, all the way from Zion Canyon to the Great Sand Dunes near Alamosa, Colo. The meteorology of the monsoon is given a few tight pages, good descriptions of the basic seasonal phenomena. (This year, by the way, saw a plentiful summer monsoon.) Nowadays weather forecasters might be a little more optimistic than they were here, and they would probably allude to the remarkable relations that have been discovered between sea temperatures off Peru and the strength of the monsoon rains across the world.

CHAINING THE HUDSON: THE FIGHT FOR THE RIVER IN THE AMERICAN REVOLUTION, by Lincoln Diamant. Carol Publishing Group, 1989 ($21.95).

At Philadelphia the King of Great Britain had just been pronounced "a Tyrant..., unfit to be the ruler of a free people." Britain was itself in the grip of another revolution, under the black banner of King Coal, that had not yet arrived in America. Steam engines in mines were nothing new, but the first steam engine to be used for any purpose beyond pumping water was set to work blowing a British blast furnace only in 1776. American ironmasters were well established at the time, their wrought iron recognized as being of high quality and fairly priced, with an annual production perhaps a third of the British total. But all iron smelting in far-off British America was still done in the old way: abundant hardwood charcoal fusing the rich black magnetite ore, the furnace bellows pumped and the heavy trip-hammer set to pounding not by new-fangled steam power but entirely by water power. This was an industrial technology successfully built of timbers and stone and skill, flourishing although soon to fade.

War and commerce alike had made it clear to the Americans that their revolutionary unity could not survive a forced partition along the Hudson River valley, which joined the ripe grain of the west to the horses and cattle of the east. The lower fjord was naturally open to sea-keeping ships as far north as Troy; it could be denied them most easily along the winding narrows of the Hudson Highlands. From 1775 until peace came in 1783, the American government undertook a series of Strategic Defense Initiatives that were intended in one way or another to close the Hudson passage against the frowning men-of-war of the Royal Navy.

The grandest accomplishment of the engineers was a wonderful chain of some 800 heavy wrought-iron links strung 1,700 feet to block the Hudson at West Point in the Highlands. Its design was made and its construction supervised by English-born Captain Thomas Machin, 32 years old, a member of the Tea Party, wounded at Bunker Hill—an officer certified by George Washington himself as "an ingenious faithful hand ... that has considerable experience as an Engineer." The 60 tons of links, each fashioned from a bar hand-forged under the blows of the water-powered tilt hammer, were floated on the river surface by a string of moored log rafts. Just downstream from that strong chain was a floating boom of iron-linked logs intended to slow the ponderous approach of an 850-ton ship of the line. The ends of the chain were well set in rock-filled cribs near the two banks, where strong forts were built and manned; the chain could be lifted by a big man-handled capstan in the fall and reset in the spring, to avoid damage by the ice that packed the wintry river.

That chain was a weapon of opportunity. For 40 years the lonely Ramapo hills along the New Jersey–New York border, only 25 miles from West Point, had been the seat of well-known iron furnaces working rich local ore. The Sterling Iron Works was awarded the chain contract and was paid a high price for speed and quality, with draft exemption assured for its furnacemen and forgemen. Work began in January of 1778, and by May 1 the chain spanned the river. Most of the furnace work was done in February and March, a ton or so of pig iron poured each day, the big stone furnace in blast around the clock. The iron bars were the yield of 100 acres or so of Ramapo hardwood forest, cut that winter to supply the four-foot chestnut billets that fed the fires. Oxcarts hauled groups of nine joined links on a valley road to New Windsor on the banks of the Hudson upstream, where they were assembled and rafted a few miles downriver to the crossing.

In the course of the five wartime summers during which it barred the way, "General Washington's Watch Chain" was never challenged by warships: perhaps it was deterrence, perhaps mere good fortune. (The treachery of General Arnold, in command at West Point, would have meant its capture, but his intrigue was laid bare in the nick of time.) As soon as peace came, the chain was sought at surplus prices by eager ironmongers nearby, but fear of another British war to come led Congress to reject the bid. The rusting links remained piled on the shore—a few were surely lost and replaced by spares in the years of handling—until the U.S. Military Academy was founded in 1802, when they were set aside, most of them to be sold in 1829 for melting into cannon rail at a foundry nearby. Today a dozen links are ceremonially mounted at Trophy Point on the Academy grounds, and a few other authentic relics of the chain are displayed here and there in the Northeast.

During the mid-19th century very heavy chain of mill-rolled iron, its links much more impressive than Captain Machin's handmade two-inch bar, was regularly produced for warship moorings by a big Welsh mill. Around 1900 a couple of New York arms dealers coolly offered for sale links of that commercial chain, under the claim that they were old pieces from West Point. The scam took in even curators at the U.S. Coast Guard Academy and the Smithsonian Institution until the 1950's; from Hollywood to Chicago you might find a few obviously spurious links still proudly labeled. For the real thing, "go to West Point."

The lively pace and dramatic style of this unexpectedly apropos study are truly enjoyable. The other projects of the ambitious SDI offices of the day—an earlier, lighter Hudson chain, a string of signal beacons, a couple of submerged log barriers and so on—were all less successful. But one of those failures was portentous. It was the first of all military submersibles:

a one-man oaken craft made by David Bushnell, an unconventional Yale student. It traveled a few feet below the surface, propelled by sealed oars, and could stay submerged "without any Inconveniency" for 45 minutes between visits to the surface for air. Two attacks on British warships anchored off Manhattan, intended to affix time bombs below their keels, were near misses, one at the foot of West 110th Street and the other near Bedloe's Island. Although Army submariner Sergeant Ezra Lee survived the risky campaign, builder Bushnell gave up the expensive effort for good.

THE WORKS OF CHARLES BABBAGE, edited by Martin Campbell-Kelly, in 11 volumes. Published in the U.S. by New York University Press, 1989 ($995).

Ten inches of blue-bound volumes, all the texts and mathematics newly and uniformly set, with title-page facsimiles and reproductions of the many original tables and line drawings, span this trim British-produced presentation of the works of Charles Babbage of London and Cambridge. The name of Babbage is widely known; he was a prophetic and yet "enigmatic figure in the history of science and of technology." Most of the material is compiled from works already printed, although some appears for the first time from manuscript papers and letters. (As yet there is no full edition of his wide correspondence.)

The largest single volume, the first one, presents Babbage's 20-odd more or less professional mathematical papers—all of them, except for one encyclopedia article completed before he turned 30, at the end of 1821. The next two volumes present his most original work, that on calculating machines: proposals, reports, comments and drawings, along with essential period accounts of the work by others, including his two sons. Here you can read the "Sketch of the Analytical Engine," with its sample program and its celebrated figure of speech: "the Analytical Engine *weaves algebraical patterns* just as the Jacquard loom weaves flowers and leaves." That sketch was published in 1843 by a young friend, the mathematician Augusta Ada, Lady Lovelace, who translated and extended it from a French paper (also given here) written by an engineer from Turin who had met Babbage and discussed his ideas.

Two more volumes show Babbage's lifelong interest in a miscellany of topics from science and technology: brief pieces on everything from lathe tools, ciphers and geology to the designing of unambiguous signals from lighthouses. Six final volumes contain editions of Babbage's six books. The first was a layman's introduction to life-insurance principles and practices; then came a polemic on the genuine decline of science in England in 1830 and another on the Exhibition of 1851, written before it opened to criticize rather sourly its location and organization. There followed a fragmentary study of the relation of science and religion and finally two substantial books, much read and much worth reading. One was a pioneer study of machines and manufacturing processes (whose value was recognized by Mill and Marx alike); the last was a crochety, vivid and instructive autobiography, published in 1864. Charles Babbage died in 1871 at the age of 79.

Although this publication does suggest that reading primary sources can be expensive, the effect of the fine compilation is no small one, even for the general reader. From the first the mind of Charles Babbage seems to have turned to the representation of information. Automata, ciphers, signal systems, printed tables of functions (and their errors) and mathematical notation (first for ordinary calculus and then for functional analysis) all plainly interested him before he began work on mechanical calculation. A second strand appears, too, in the mathematical models he devised for everyday matters, closer to operational analysis than to mathematical economics, but touching both.

Babbage is universally regarded as a pioneer of the programmed computer, but that is not because his early and ill-recorded work had any demonstrated direct influence on computer development and still less because it was influential in his own time. Rather, it is the recognizably modern pattern of his thought—a kind of topic outline of the interests in and around computer laboratories today—that catches his modern peers. His was an elaborated idea, offering historians the rare test case of the "development of an idea in two independent contexts," as computer historian Alan G. Bromley comments. By 1840 Babbage had recognized the need for a memory and a central-processing unit and for cards to hold both entry data and sequential programmed operations that could function conditionally, their subroutines held on pin cylinders like a music box. His program was not modifiable except by the user, Bromley points out.

Why did the general-purpose computer fail the first time around? We know some reasons. It failed for lack of conspicuous applications for an expensive analytical engine. (Printed books of mathematical tables, Babbage's prototype of what his engine could economically do, have ironically been all but outmoded by the machine's own quick-calculating successors.) It failed because so complex a machine could simply not be realized in tons of interconnected brass and pewter clockwork. And to a degree it failed as well because of the impolitic and combative style of its spirited promoter. A fuller story needs still to be worked out.

THE NEW PHYSICS, edited by Paul Davies. Cambridge University Press, 1989 ($49.50).

Gold is where you find it, the prospectors say. Paul Davies, theorist and successful popular expositor of physics at the University of Newcastle upon Tyne, has induced a starry Anglo-American cast (and one author from Brussels and another who splits time between London and Trieste) to expound the "new physics," which he sees as inheritor to the Golden Age, the time of quantum fields, neutron and positron. There is surely the glitter of gold here, particularly in the unifying work of the 1970's. But what we read is a little narrower than the claim; it is not all of physics but rather theoretical physics that is front stage. At most one or two of the authors would claim to be at home in any laboratory. The continuity of theoretical ideas is so strong—much stronger than it was across the years from 1924 to 1932—that progress, while steady, is not always shiny new. If you extend the turn-of-the-century studies of Henri Poincaré on complex orbits but yet manage to omit his name, your beautiful computer physics may look a little newer than it is.

The essays that appear in this hefty volume divide into five topics, plus a brief summary by Davies, who posts three frontiers within physics: the small, the large and the complex. The longest of the book's five sections deals with the large: the "new astrophysics" and the cosmology that goes with it. A second section is of course that of the particle theorists, ending with a brief, personal and provocative piece by Nobel laureate Abdus Salam. Another group of papers is concerned with the physics of condensed matter, surfaces, phase changes that generalize the orderly surprises in superconductivity and a deft mixture of atomic and photon physics called quantum optics. A fourth section centers on the new issues of statistical behavior, both the remarkable self-organizing nature of open systems as energy flows through them and the quasi-indeterminacy now called chaos. Finally, there is a single penetrating if not at all elementary article by Abner Shimony on the conceptual foundations of quantum mechanics itself, the modern chase after a hare started by Einstein and Bohr in the 1930's.

What a reader meets here is measurable in units of *Scientific American* articles. Deploying a similar, although rather less elaborate, display of graphic aids and a considerably higher content of mathematics, the hefty volume runs to 18 essays, about 500 pages. Five of the briefer essays are each like a single magazine article; a dozen more are double-length articles on a single topic. There is too much variety to sample satisfactorily.

That longest single contribution, by Malcolm S. Longair, amounts to a lively brief book on current astrophysics. Longair is one who gives observation and experiment their due; he sees most of the theoretical developments (apart from black-hole physics not yet verified) as being largely stimulated by observational findings. Clifford M. Will recounts the postwar renaissance of general relativity in the same vein, and other pieces treat the difficulties of quantum gravity and the promise of inflationary cosmology. The greatest change since the Golden Age (in 1936 Bohr had prophesized that it would come) is the evidence that within one single physical context—the early universe, in part open to our inquiry—very large and very small are jointly relevant.

Quarks and the gauge-field theories that include and go beyond that level of analysis of matter are of course prominent. A personal story of how "grand unified theories" grew in the 1970's from simple group to group is told most ingratiatingly by Howard M. Georgi, even admitting to his present doubts about strings. On the other hand, Professor Salam is hopeful of recent work with 26 dimensions, which can be made equivalent to 26 fundamental fields in a two-dimensional spacetime. The "incandescent beauty of this theory" manifests itself in the uniqueness of certain (untestable?) predictions. Here what is most novel is the decisive importance of modern mathematics—topology, for example, once thought to be of little moment to the physicist. "One can only hope nature is aware of our work on all this," Salam says wistfully.

ESSAY

How to lose: the story of maglev

by Daniel Patrick Moynihan

This is the story of a contest almost no one is watching. At stake is preeminence in the production and sale of a revolutionary new mode of transportation. It is called magnetic levitation—maglev for short. It will define the coming century much as the railroad defined the last one and the automobile and airplane have defined this one.

Maglev is a combination railroad, automobile and airplane. A maglev vehicle is essentially a fuselage without wings that travels along a magnetized guideway at speeds up to 300 miles per hour. The principles behind it are well established.

The idea for maglev was thought up by Dr. James R. Powell on a Friday night in February, 1960, while he was stuck in that permanent traffic jam that awaits anyone trying to leave Long Island over the Bronx-Whitestone Bridge. Powell, a 28-year-old nuclear engineer at the Brookhaven National Laboratory on Long Island, figured you could exploit magnetism to both levitate and propel a vehicle along a guideway. Powell and Dr. Gordon T. Danby, a colleague at Brookhaven, presented the first paper on superconducting maglev transportation at an engineering conference in 1966, and the race was on.

At the start, of course, the U.S. was the only entrant. Development work began at M.I.T. under the guidance of Dr. Henry H. Kolm. By 1974 Kolm had developed a 1/25th scale model of a maglev vehicle he dubbed the magneplane. It moved millimeters above its aluminum track at 56 miles per hour. The federal government saw some promise in the idea. The High Speed Ground Transportation Act of 1965 provided funds for research and development of high-speed trains, air-cushion vehicles and, later on, maglev. Grants were awarded to Ford Aerospace and the Stanford Research Institute. We were on our way.

Then everything stopped. On February 5, 1975, a representative of the Office of Management and Budget announced that maglev research was no longer a priority. The budget would be juggled to shore up the past century's conventional rail technology. But by this time we were not alone in the race. Other countries had been quick to pick up on Powell and Danby's idea. By the late 1960's Great Britain, Japan, West Germany and—can you imagine?—Romania had initiated their own maglev programs. Japan and Germany have stuck at it.

By 1972 the Japanese had a model going; by 1979 they had tested a maglev vehicle at 320 miles per hour. The full-size German Transrapid 06 has traveled 100,000 miles on its 20-mile test track in Emsland. And these projects did not just happen. The Japanese government has poured $1 billion into maglev research, the West German government almost as much. By comparison, the U.S. government spent $3 million on maglev between 1966 and 1975; after that, nothing. Now the new secretary of transportation, Samuel K. Skinner, has said he is preparing a "national transportation plan" to be released early in 1990. If the U.S. is to make a commitment to maglev, this is the place to do it.

During its short history this country has experienced several transportation revolutions. Each of them has had a profound and irreversible effect on society.

The opening of the Erie Canal in 1825 began the great era of waterborne commerce. Canals provided convenient access to the interior and could carry cargo at a fraction of the cost of the horse and wagon. Thenceforth the federal government contributed substantially to the construction of the canal network.

Inexorably, canals were superseded by railroads, although not without some complaint. In 1829 New York's Governor Martin Van Buren wrote a strongly worded letter to President Andrew Jackson:

> As you may well know, Mr. President, "railroad" carriages are pulled at the enormous speed of 15 miles per hour by "engines," which in addition to endangering life and limb of passengers, roar and snort their way through the countryside, setting fire to the crops, scaring the livestock and frightening women and children. The Almighty certainly never intended that people should travel at such breakneck speed.

Nonetheless, by 1860 the canals were on the wane and the railroad was ascendant—again with federal help, this time in the form of massive land grants in the West.

Since that time the railroad has changed very little. Just last fall Amtrak proudly announced that its Metroliner had averaged 110 miles per hour on a special Boston to New York run—two miles per hour slower than the mark set by the Empire State Express No. 999 between Batavia and Buffalo in 1893! The Japanese Bullet train and the French TGV do a little better (operating speeds range from 150 to 180 miles per hour), but beyond this they will not advance. They have reached the upper limit for wheel-on-rail technology.

In the early years of this century the automobile quickly overwhelmed the railroad. By 1918 there were 6.2 million cars in the U.S. In what was perhaps the last great moment for the automobile, Congress enacted a highway bill in early 1987 to complete the Interstate Highway System. The Interstates, begun in 1956, have made up the largest public-works project ever undertaken, and their completion in the early 1990's will mark the beginning of the end of the Automobile Age. The Interstates, while still useful, will have become an idea of the past.

As for the airplane, it has served us well, but the air transportation system is already operating beyond the reasonable limits of its capacity. We can build more airplanes, and we can even build more airports (although we do not seem to be very good at this: no new major airport has opened since Dallas/Fort Worth did in 1974, even though passenger traffic has doubled). But try as we might, we cannot build more airspace. A finite amount of it is available over any given airport, and the airspace near many major cities is full to capacity.

The long and short of it is this: no major transportation revolution has ever progressed in the U.S. without substantial federal involvement. Maglev will be no exception. And if we do not build our own maglev system, we will end up buying it from others.

Fifteen years from now, when you settle into your seat for the one-hour run from Chicago to Detroit, have a look under your seat cushion. My sincere hope is that you will not find a small tag that reads "Maglev: invented by American scientists, made in West Germany."

DANIEL PATRICK MOYNIHAN is the senior U.S. senator from New York.

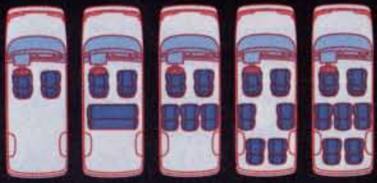
Lumina APV. What

The future holds what no one has ever seen before. A vehicle so unique it doesn't have a look-alike. It has a body that will never rust. Available modular bucket seats that can be used as seats, tables or removed. Innovations that put you first. And a design so advanced, it can take you into the future. Today.

It's First.

Nobody's offered you all these options before. To seat up to seven people, the way you want. In one standard or four optional seating arrangements.

And even the optional modular bucket seats are new. They can fold down and become tables or be removed altogether, to carry 129 cu. ft. of cargo.



Another first: standard Scotchgard™ Fabric Protector on Lumina's seats and doors. So they stay fresh-looking longer. And another: a large solar windshield designed to reflect the sun's rays. Its benefits are clearly visible.

And It Lasts.

Lumina APV's Bumper to Bumper Plus Warranty covers everything for 3 years or 50,000 miles.* In other words: it's built to last.

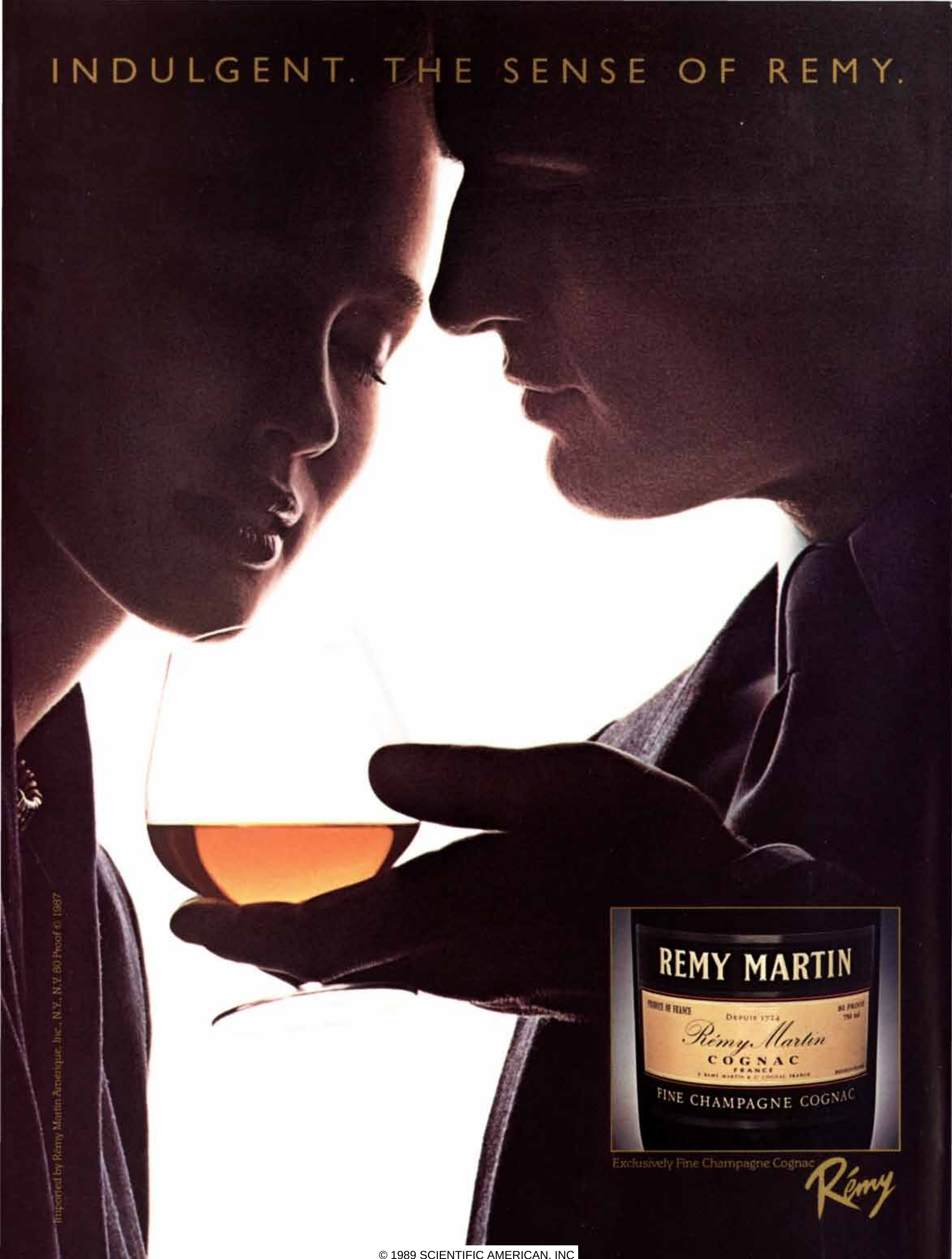
It's Unrustable.

APV's body panels are made of a new high-strength, reinforced, advanced composite material that will never rust. Resists dents. And, thanks to clear-coat paint, they'll stay bright longer.

*Except for tires which are covered by their mfr. A deductible will apply after 12 months/12,000 miles. See your Chevrolet dealer for terms of this limited warranty.

Let's get it together…buckle up.

LUMINA APV


INDULGENT. THE SENSE OF REMY.
REMY MARTIN
Depuis 1724
Rémy Martin
COGNAC
FRANCE
FINE CHAMPAGNE COGNAC
Exclusively Fine Champagne Cognac
Rémy
Imported by Rémy Martin Amerique, Inc., N.Y., N.Y. 80 Proof © 1987