SCIENTIFIC AMERICAN

JUNE 1989
$2.95

M.I.T.'s R_x for building a new industrial America.

Strange fossil creatures from an ancient sea.

How a crystal lattice channels electrons and positrons.

***An earthquake** waiting to happen? This sharply folded terrain in the foothills of the Andes could conceal a dangerous fault.*

THE SWIFT REFLEXES OF EAGLES.

PRESENTING EAGLE PREMIER.

A suspension system can be built in one of two ways. The first way isolates you from the road. The second, the Eagle way, puts you not only in touch with, but also in charge of, the road and its changing surfaces. Its twists and turns.

That's why the Premier ES has independent suspension at all four wheels. And an advanced torsion-bar rear suspension. All for superior ride and handling.

Inside this beautifully aerodynamic automobile, you'll find spaciousness. Instruments and controls positioned for the driver's advantage. And you can select genuine fine leather-trimmed seating on the Premier ES.

To take the driver's confidence a step further, every Eagle Premier is protected by Chrysler's exclusive 7-year/70,000-mile Protection Plan.*

Take a test drive. We believe you'll find Eagle Premier is a world-class way to fly.

7/70

For further information, call 1-800-JEEP-EAGLE.
Call this same number to find out about the Eagle Information/Game PC Disk.

*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.

Buckle up for safety.

Jeep
Eagle

Expect the Best.

SCIENTIFIC AMERICAN

June 1989 Volume 260 Number 6

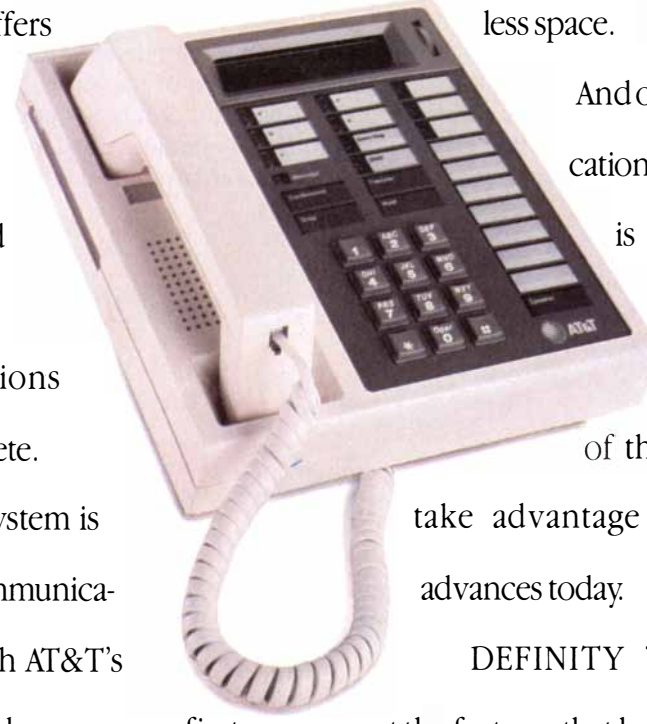
DEPARTMENTS

Hello D
goodbye ob

Imagine a communications system designed to expand right along with your business, no matter how large you grow or how complex you get. One that offers seamless communication from 40 to 30,000 lines and beyond. And allows you to add lines economically.

Imagine a communications system that may never be obsolete.

The DEFINITY™ 75/85 system is simply the most advanced communications system you can own. With AT&T's new "universal module" hardware, your first module functions as the system's base. After that, adding lines is simply a matter of adding modules.



AT&T's new high-density circuit packs allow you to add more phones and computers to your system, while requiring less power and less space.

And of course the communication system of the future is fully compatible with the telecommunication technology of the future, so you can take advantage of emerging ISDN advances today.

DEFINITY 75/85 systems also support the features that have always made AT&T communications systems so popular. AUDIX voice mail, for example, offers you twenty-

EFINITY, solescence.

four-hour access for retrieving and leaving messages. And AT&T's System Management offers you a range of tools, including windowing capabilities, to administer and maintain your DEFINITY 75/85 system.

All of which add up to increased business productivity.

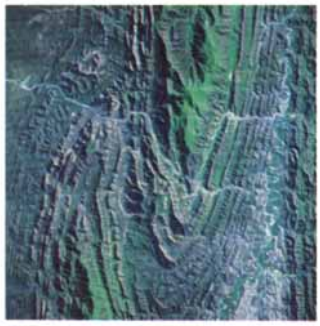
With a DEFINITY 75/85 system you've got the very best in investment protection. And if you're a current System 85 or 75 customer, you can upgrade to the DEFINITY 75/85 system while protecting up to 90% of your installed communications investment.

The DEFINITY 75/85 system was designed to evolve as your business requires. With it there'll be virtually no limit on how large you can grow and still communicate effectively.

To find out more about the DEFINITY 75/85 system, contact your AT&T Account Executive, or call 1 800 247-1212, Ext. 190.

From equipment to networking, from computers to communications, AT&T is the right choice.

AT&T
The right choice.



THE COVER photograph, made from data collected by *Landsat*'s thematic mapper and color-enhanced by the Cornell Andes project, shows highly folded terrain in the Bolivian sub-Andean belt (see "Hidden Earthquakes," by Ross S. Stein and Robert S. Yeats, page 48). This seismically active region near the Bolivian-Argentine border has not seen large earthquakes. Yet the growing folds may conceal hidden faults and are therefore potential sites for major earthquakes.

THE ILLUSTRATIONS

Cover image by Richard Allmendinger

Page	Source
41	Courtesy of Ampex Corporation
42	Johnny Johnson
43	Andrew Christie
44	Roderick Chen, Superstock
45–46	Andrew Christie
48–49	John E. Estrem, Ross S. Stein, Jerry P. Eaton and Raymond M. Batson
50	Ross S. Stein (*top*), Hank Iken (*bottom*)
51–54	Hank Iken
55	John E. Estrem and Ross S. Stein
56–57	Timothy B. Smith, Robert V. Sharp, Roger D. Borcherdt, David W. Simpson and Ross S. Stein
61	Shirley M. Taylor, Lesley A. Michalowsky and Peter A. Jones, University of Southern California School of Medicine
62–68	George V. Kelvin
74–75	Paleontological Institute and Museum at the University of Zurich
76	Tom Prentiss
77	Paleontological Institute and Museum, Zurich (*top*), Tom Prentiss (*bottom*)
78–79	Paleontological Institute and Museum, Zurich
80	Paleontological Institute and Museum, Zurich (*top*), Tom Prentiss (*bottom*)
81	Tom Prentiss (*top and bottom*), Paleontological Institute and Museum, Zurich (*middle*)
87	Quesada/Burke (*top*), Thomas C. Moore (*bottom*)
88–89	Thomas C. Moore
92	Quesada/Burke
94	Thomas C. Moore
95	Michael G. Rossmann and Sharon S. Wilder, Purdue University
97	Allan H. Sørensen
98–101	Ian Worpole
105	Stanley V. Margolis
106	J. Paul Getty Museum
107	George Retseck
108	Stanley V. Margolis
109	George Retseck
110	J. Paul Getty Museum
113	The Cleveland Museum of Art, Hinman B. Hurlbut Collection
114	Wilfred Niels Arnold, courtesy of Linda Hall Library, Kansas City, Missouri
115	Wilfred Niels Arnold and Louise M. Arnold (*top*), The Seagram Museum (*bottom*)
116	Johnny Johnson
117	S. Varnedoe
118–121	Michael Goodman
122–124	Edward Bell
130	Augusto Gansser, from *Himalayas,* © 1987 Facts on File Publications

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 $24, two years $45, three years $60 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631/32. 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, *Associate Editor;* Timothy M. Beardsley; John M. Benditt; Laurie Burnham; Elizabeth Corcoran; Gregory R. Greenwell; John Horgan; June Kinoshita; Philip Morrison, *Book Editor;* Tony Rothman; 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;* M. Knight; 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; Julio E. Xavier

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. 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. ATLANTA: Quenzer/Stites. CANADA: Fenn Company, Inc. DALLAS: Griffith Group. PRINCETON: William Lieberman, Inc.

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.

ASSOCIATE PUBLISHER/BUSINESS MANAGER: John J. Moeling, Jr.

PRESIDENT OF MAGAZINE DIVISION AND PUBLISHER: Harry Myers

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; *Senior Vice-President,* Harry Myers; *Vice-Presidents:* Linda Chaput, Jonathan Piel, Carol Snow

CHAIRMAN OF THE BOARD: Georg-Dieter von Holtzbrinck

CHAIRMAN EMERITUS: Gerard Piel

THE BASIC IDEA BEHIND THE VOLVO TURBO WAGON.

Basically, the rig pictured above can do anything a Volvo 740 Turbo Wagon can do.

It will hold loads of cargo. Out-accelerate many performance cars. And reach top speeds of over 100 MPH. It will also pull about 150 G's.

Right from your wallet.

Which is why it pays to consider a Volvo 740 Turbo Wagon instead.

Its 74 cubic feet of cargo space is greater than most wagons in its class. While its performance rivals that of some of the world's top sports cars.

You see, unlike other wagons, *both* ends of the Volvo 740 Turbo Wagon were built for hauling.

VOLVO
A car you can believe in.

SCIENTIFIC AMERICAN

SPECIAL Single-Topic Issue

MANAGING PLANET

September 1989

Each year, SCIENTIFIC AMERICAN devotes an entire issue to an in-depth look at a single topic. The single-topic issue has focused on such important subjects as the brain, advanced computing, economic development and AIDS. This year's topic, Managing Planet Earth, marks the 40th consecutive SCIENTIFIC AMERICAN single-topic issue. And it may well be the most important yet.

"As the 20th century draws to a close, one issue looms as the central challenge that we face: Can the earth be made safe for human beings and other forms of life?

"Calling a worldwide industrial and agricultural time-out is not a real option. Neither is journalistic hand-wringing that fails to offer solutions. Somehow humankind must devise a strategy that reconciles economic growth and development with a habitable environment.

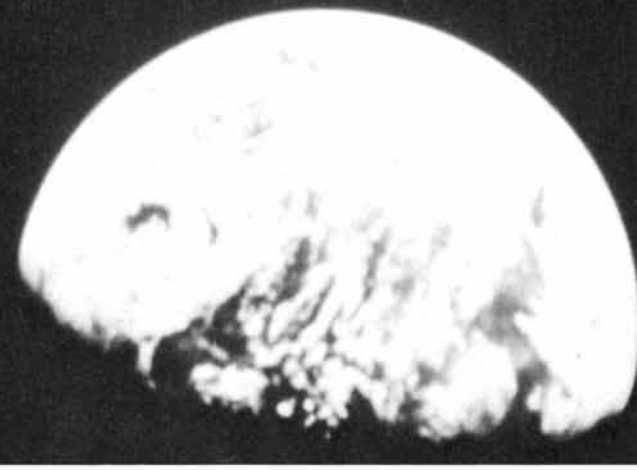
"The Editors of SCIENTIFIC AMERICAN have invited the most eminent biologists and earth scientists to join with the most knowledgeable and creative industrial executives and concerned political leaders to map out a bold, effective strategy for meeting the challenge of managing planet Earth."

Jonathan Piel, Editor

Some of the subjects to be covered in this SPECIAL ISSUE

- **GLOBAL CHANGE—Introductory article**
- **THE CHANGING ATMOSPHERE**
- **THE CHANGING CLIMATE**
- **THE CHANGING HYDROSPHERE**
- **THE CHANGING BIOTA**
- **AGRICULTURE IN A CHANGING ENVIRONMENT**
- **ENERGY USE IN A CHANGING ENVIRONMENT**
- **MANUFACTURING IN A CHANGING ENVIRONMENT**
- **ECONOMIC DEVELOPMENT IN A CHANGING ENVIRONMENT**
- **MANAGING THE WORLD'S POPULATION**
- **POLICY PATHS FOR THE FUTURE**
- **ESSAY: OUR COMMON FUTURE**

EARTH



SPECIAL ADVANCE ANNOUNCEMENT

Due to the expected high interest in this September 1989 single-topic issue and a limited print run, the publisher is offering a special opportunity to our readers to order extra or bulk copies in advance.

"Managing Planet Earth" will be a path-breaking issue that belongs on everyone's reading list and should be included in next year's curriculum. Share this vital information with your colleagues, students and anyone else for whom these pressing issues are a concern. Clip the Pass-Along Order Form below and bring it to their attention!

ORDER FORM

Yes, enter my order for _____copies of the September 1989 single-topic issue, "Managing Planet Earth", at $_____ per copy. Total order $_____.

Bulk and prepaid orders receive substantial discounts! Quantities limited. Order by June 30, 1989.

*1 to 9 issues $2.95 per issue *25 to 99 issues $2.50 per issue
*10 to 24 issues $2.75 per issue *100 or more issues $2.20 per issue

NOTE Deduct 10% from total with paid order.

PRICES INCLUDE SHIPPING (U.S. ONLY)*

☐ $_____ Payment enclosed ☐ Purchase order attached
☐ Charge my
☐ VISA ☐ MasterCard ☐ Access ☐ Eurocard ($10 minimum)

Card # _____ Exp. Date _____

Signature _____

Name _____ School _____

Address _____

City _____ State _____ Zip _____

For more information, call Ms. Davis at 1-212-754-0539.
Expected shipment date Aug. 29, 1989.
Cut and mail this coupon to: 415 Madison Avenue, Dept. MPE, NY NY 10017
*Outside the U.S. add $1.00 per copy.

P9AA110

PASS-ALONG ORDER FORM

Yes, enter my order for _____copies of the September 1989 single-topic issue, "Managing Planet Earth", at $_____ per copy. Total order $_____.

Bulk and prepaid orders receive substantial discounts! Quantities limited. Order by June 30, 1989.

*1 to 9 issues $2.95 per issue *25 to 99 issues $2.50 per issue
*10 to 24 issues $2.75 per issue *100 or more issues $2.20 per issue

NOTE Deduct 10% from total with paid order.

PRICES INCLUDE SHIPPING (U.S. ONLY)*

☐ $_____ Payment enclosed ☐ Purchase order attached
☐ Charge my
☐ VISA ☐ MasterCard ☐ Access ☐ Eurocard ($10 minimum)

Card # _____ Exp. Date _____

Signature _____

Name _____ School _____

Address _____

City _____ State _____ Zip _____

For more information, call Ms. Davis at 1-212-754-0539.
Expected shipment date Aug. 29, 1989.
Cut and mail this coupon to: 415 Madison Avenue, Dept. MPE, NY NY 10017
*Outside the U.S. add $1.00 per copy.

P9AA128

LETTERS

To the Editors:

In "AIDS in 1988," by Robert C. Gallo and Luc Montagnier [SCIENTIFIC AMERICAN, October, 1988], the authors suggest that AIDS had been present in central Africa for many years before the current epidemic. To support this theory they argue that the central Africans may have had little contact with the outside world, thereby containing the disease.

The authors do not seem to be aware of central African demography and history. Communication between city and country and migration in both directions have been part of central African society at least since World War II. The health services in Uganda, the region some claim may be the source of HIV, the AIDS virus, were very good in the 1950's, 1960's and 1970's. AIDS would have been identified if it had existed. I went to medical school and practiced medicine in Uganda in the 1970's, and I can assure Drs. Gallo and Montagnier that there were no cases of AIDS or AIDS-like diseases at that time. The first cases in Uganda were seen in 1982, after the first American cases had been reported.

In a study published in the *British Medical Journal* (September 27, 1986), blood samples from 6,015 Africans were examined for antibodies to the AIDS virus. These samples had been collected between 1976 and 1984 in nine western, central and eastern African countries, including Uganda. Only four contained antibodies! The investigators concluded that fewer than one in 1,000 subjects were seropositive for AIDS at the time of sampling. They discounted the hypothesis that the disease originated in Africa. I note that this study is never referred to by the "experts" or by the media.

The fact is that we do not know the origin of HIV. Of course, it is very important that we identify its source in order to learn more about HIV and other viruses. But perpetuating a myth does not help science and only reinforces the ever-present racism and bigotry toward Africans.

MUNIINI MULERA

Department of Pediatrics
University of Toronto, Ontario

The authors respond:

Dr. Mulera raises several important issues about our interpretation of the data regarding the origin of the AIDS virus. We stated in our article that the virus is a naturally occurring organism and postulated that it has been present in "small, isolated groups in central Africa or elsewhere for many years." We do know that the epidemic began at approximately the same time in the U.S., Haiti and central Africa; that many species of African monkeys are infected with related retroviruses; that these simian retroviruses rarely, if ever, cause disease in their specific host, which suggests each simian species has had a long time to adapt to its own virus; and that, according to some estimates, HIV-1 has been in humans more than 100 years, although it does not appear to be an ancient infection. We and others have suggested that demographic changes may have resulted in wide dissemination of the virus from smaller, closed groups into larger, susceptible groups, creating new patterns of disease.

There are several reasons to suggest that the virus has been present and causing disease in areas of Africa longer than anywhere else. Although the clinical description was first made in the U.S. in 1981, cases of AIDS have been identified in retrospect in central Africa as early as the 1960's (for example, see "Early AIDS Cases Origi-

nating in Zaïre and Burundi (1962–1976)," in the *Scandinavian Journal of Infectious Diseases,* Vol. 19, pages 511–517, 1987). The earliest known positive sera are found in African samples from the 1950's (see "Evidence for Human Infection with an HTLV III/LAV-like Virus in Central Africa, 1959," in the *Lancet,* Vol. 1, pages 1279–1280, May 31, 1986). Serology in a Zairian village revealed a 1 percent rate of infection that remained unchanged from 1976 through 1986, which also suggests an established disease. High rates of infection in certain areas (in 1983 it was found in 80 percent of prostitutes in Rwanda) suggest the more recent introduction of virus into groups with high rates of transmission and highlight the magnitude of the epidemic in comparison with earlier sporadic cases. In contrast, the study cited by Dr. Mulera reviewed samples largely from western Africa.

Although human retroviruses are found in other areas and populations, there are several reasons those of us interested in human retrovirology have an interest in Africa. There is a connection between the known human retroviruses (including HTLV-I) and related simian viruses found in wild African monkeys but never in New World primates. Although retroviruses closely related to HTLV-I are also found in Asian primates, they are not as closely related to the human virus as the African viruses are. Simian retroviruses related to HIV have not been found in either Asian or American primates. Although we have not found a very close simian relative of HIV-1 in Africa, some simian immunodeficiency viruses (SIV's) are about 50 percent related to HIV-2, another human retrovirus. HIV-2 is remarkably restricted to parts of western Africa. Earlier sera (from the 1970's) have been positive only from that region, and so we conclude that HIV-2 probably has been present there for a long time.

The origin of the AIDS virus is obviously related to but different from the factors that govern the epidemic in each group or area. We do not suggest that the virus was transmitted en masse from central Africa to other areas or that the people of that region are any more responsible for the disease's transmission than, say, the residents of Connecticut are responsible for Lyme disease, which in fact was first described in Lyme, Conn. What we postulate as the trigger of the epidemic in central Africa—changing demography and behavior—is equally applicable to groups in Europe and the Americas. Historically we note that many microbes appear to have originated in Europe and the Americas or were first found in humans there.

We deeply regret, and in fact do not understand, an interpretation of our work and opinions on the origin of HIV that finds bigotry or suggests we blame anyone, much less an entire continent. In our opinion, HIV-1 very likely originated in a small region of central Africa and HIV-2 in western Africa. This point is not based on prejudice, nor is it purely academic; it has major long-term medical importance, and we believe it is time to discuss facts openly and honestly.

ROBERT C. GALLO

National Cancer Institute
Bethesda, Md.

LUC MONTAGNIER

Pasteur Institute
Paris

To the Editors:

I was intrigued by the behavioral therapy for obsessive-compulsive disorders described in "The Biology of Obsessions and Compulsions," by Judith L. Rapoport [SCIENTIFIC AMERI-

CAN, March]. As a student of the history of the Roman Catholic Church, I found the treatment very similar to what is advised in confessors' manuals, beginning at least as early as the Council of Trent in the 16th century. Religious scrupulosity became a particular problem for the Church at that time because of the rapid growth of the Jansenist heresy. The rigorist moral code of the Jansenists was very attractive to certain poor souls. Parish priests might be confronted in the confessional by a man who felt it was a sin not to raise his hat when a priest walked by, for example, or a woman who feared the Divine Wrath if she failed to light a certain number of candles.

The manuals advised the priest to assure the penitents that these were not sins and then to forbid them specifically from performing these actions. The man would be told (under pain of sin) *not* to raise his hat and the woman *not* to light any candles in church. My impression is that the method had at least some effect.

Ordinary Catholics were encouraged to select a personal confessor, to whom they would always go for advice and absolution—in part to try to head off the development of unhealthy religious habits. The biographies of many of the saints catalogue the effects of these practices from both sides of the grille; see, for example, the lives of St. John Vianney and St. Dominic Savio.

THOMAS P. BYRNE

Notre Dame High School
Belmont, Calif.

The author responds:

Mr. Byrne makes a good point. Interested readers might refer to my book, *The Boy Who Couldn't Stop Washing: The Experience and Treatment of Obsessive-Compulsive Disorder* (E. P. Dutton, 1989). It has an appendix describing the Church's long and enlightened efforts to treat such scrupulosity.

JUDITH L. RAPOPORT

National Institute of Mental Health
Bethesda, Md.

To the Editors:

Our recent article, "Soft-X-Ray Lasers" [SCIENTIFIC AMERICAN, December, 1988], did not provide sufficient in-depth coverage of the wide variety of work of our many colleagues. Space limitations together with the constraints involved in writing an article for nonspecialist readers made it impossible to discuss all the many kinds of research (including many aspects of our own work) as thoroughly as they deserved.

For the record we would like to mention the major international research efforts toward developing X-ray lasers taking place in France (University of Paris and the French Atomic Energy Commission), England (Rutherford Appleton Laboratory and University of Hull), West Germany (Hahn-Meitner Institute at Garching), Japan (University of Tokyo and University of Osaka), Canada (the National Research Council in Ottawa), the Soviet Union, China, Israel and East Germany. In this country there are important programs similar to ours, namely at KMS Fusion, Inc., in Ann Arbor, Mich., the Massachusetts Institute of Technology, the Naval Research Laboratory and Princeton University. Interesting new approaches using ultra-short laser pulses (less than 10^{-12} second) and extremely powerful (more than one terrawatt) optical lasers are being pursued at some of the above institutions as well as at the University of California at Berkeley,

the University of Chicago and Stanford University. Also of note is the work being done with large pulsed-power machines at Sandia National Laboratories, the Naval Research Laboratory and Physics International Company in San Leandro, Calif.

First claims of demonstrations of X-ray lasers date from the mid-1970's and early 1980's, with the efforts of Pierre Jaeglé and his co-workers at the University of Paris, A. Ilyukhin, I. Sobel'man and their colleagues at the P. N. Lebedev Physical Institute and Geoff Pert and his co-workers at the University of Hull. Unambiguous demonstrations of substantial amplification came later from the Lawrence Livermore National Laboratory (*Physical Review Letters,* Vol. 54, pages 106–109; January 14, 1985) and from Princeton (*Physical Review Letters,* Vol. 55, pages 1753–1756; October 21, 1985). Since that time many new reports have surfaced from all the institutions mentioned above.

Significant applications of the X-ray laser other than the holography we mentioned in the article are currently being explored at Princeton (X-ray microscopy of cells in vitro) and at Livermore in collaboration with Obert R. Wood II and William T. Silfvast of the AT&T Bell Laboratories (photoionization lasers pumped by a focused X-ray laser). The development of X-ray–laser double-pass amplifiers with a multilayer X-ray mirror at one end of the amplifier has also been proceeding at Princeton.

Of the work done at our own laboratory, the article did not mention the early pioneering efforts of George F. Chapline, Jr., Peter L. Hagelstein, Thomas A. Weaver and Lowell T. Wood, all of the Lawrence Livermore National Laboratory. In addition, it did not discuss the importance of beam refraction and the role our exploding-foil amplifier plays in alleviating it, even though the foil amplifier has produced the majority of today's powerful X-ray lasers because it can minimize beam steering within the amplifier. We were also remiss in having omitted mention of the excellent target-fabrication work done in the laser program at Livermore and at the LUXEL Corporation in Friday Harbor, Wash. KMS Fusion, Inc., also helped us with crucial experiments that paved the way for demonstrations of X-ray amplifiers. Finally, credit is due the many people at Livermore for the sophisticated atomic data base and large computer simulation codes on which all of our predictive capability rests.

We would like to correct a couple of errors in the manuscript. First, when we were describing the pioneering mirrors of Eberhard Spiller at IBM's Thomas J. Watson Research Center, we were speaking of multilayer X-ray mirrors, not grazing-incidence reflection optics such as had been developed much earlier by P. Kirkpatrick and Albert V. Baez. Second, we produced the first hologram *made by an X-ray laser,* rather than the first hologram made by X rays.

DENNIS L. MATTHEWS

MORDECAI D. ROSEN

Lawrence Livermore
National Laboratory
Livermore, Calif.

EDITOR'S NOTE

Michael D. Norenberg, co-author of "Astrocytes" [SCIENTIFIC AMERICAN, April], would like to point out that he is currently on the staff of the Veterans Administration Medical Center in Miami, Fla. His research has been supported by the Veterans Administration for many years.

It's not how fast you go.

GLI 16V At Volkswagen, we engineer cars for the German Autobahn.

Where there are two speeds.

Fast. And faster.

Here, cars must handle with pinpoint precision. Provide better directional control.

And a greater sense of road holding.

Everything a Volkswagen does naturally.

So, it's not surprising that Volkswagen is the most popular car on the Autobahn.

Consider our Jetta GLI 16-valve.

"True to its German heritage, it feels completely secure at speed." (*Car and Driver*) We couldn't agree more.

For details on the 1989 Volkswagens call 1-800-444-VWUS.

It's how well you go fast.

Of course, we don't recommend you drive at Autobahn speed.

Point is, if a Volkswagen can make you feel at ease traveling upwards of 100 miles per hour, imagine how comfortable it can make you feel at 55.

So may we suggest you see your Volkswagen dealer for a test drive.

Lots of performance cars are capable of going very fast. A Volkswagen is one that deserves to.

German engineering. The Volkswagen way.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JUNE, 1939: "For all the prodigious efforts spent in the attempt to discover drugs having marked action in germ diseases, today there are only a few great specifics, such as quinine for malaria and the arsenical for syphilis. Within less than four years, two chemical agents of outstanding value to medicine have been introduced. One, sulfanilamide, has gained established application in treating, often curing, a number of acute infections caused by particular strains of bacteria. In particular, it has revolutionized the treatment of gonorrhea. The other, sulfapyridine, already appears to rival, if not surpass, sulfanilamide's established values."

"The $50,000,000 annual product value of chewing gum in the U.S., divided by five cents (the price of a packet), represents roughly a billion purchases containing about ten billion chews. Allowing three hours per chew, and assuming the brake horsepower of the prime mover at, say, one thirtieth (although some we see look more like a full horsepower and often sound like ten), we arrive at something like one billion horsepower-hours of power expended the nation over, each year, in chewing gum."

"For the source of stellar energy, we must look to some process that changes the mass of atoms and liberates a corresponding amount of energy—in particular, to reactions between charged atomic nuclei. Here a notably successful theory has been developed by Professor Bethe of Cornell, along lines resulting from the recent work of Gamow and others."

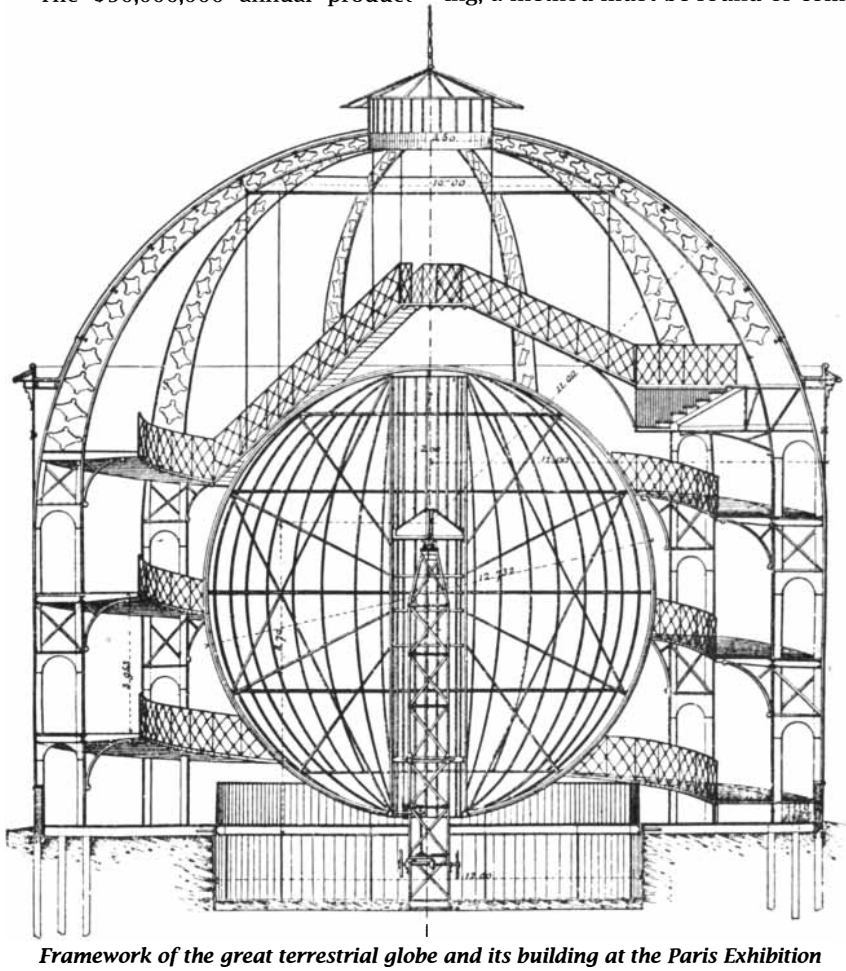
SCIENTIFIC AMERICAN

JUNE, 1889: "Before water gas can be commonly used for heating or lighting, a method must be found of communicating to this gas, which is in itself inodorous, a scent much more penetrating and persistent than that of coal gas, since, by reason of the large proportion of carbon monoxide which it contains, it is at least five times more poisonous than coal gas. During the months of January and February, 1888, seven cases of poisoning from water gas have occurred in New York."

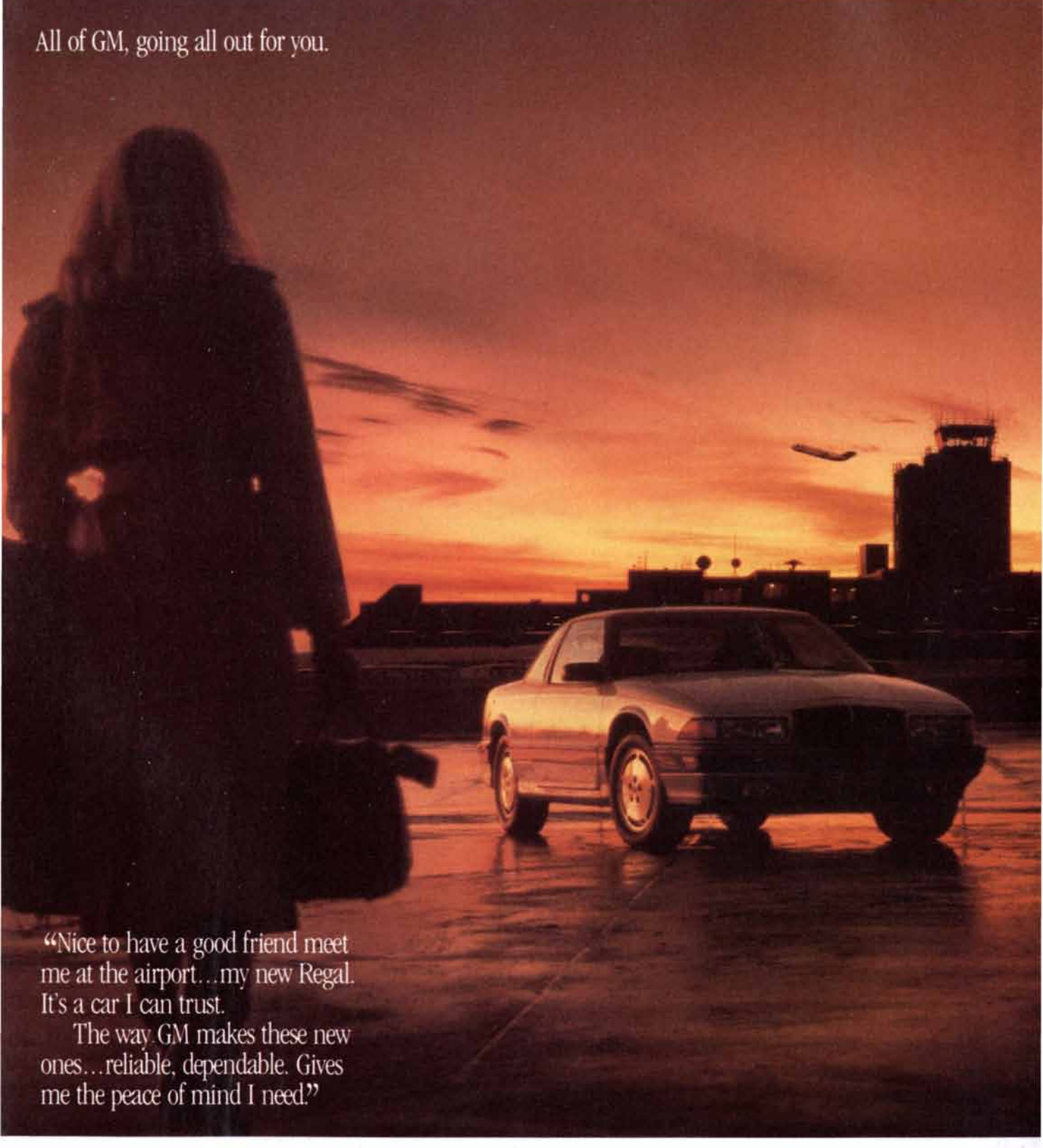
"A new and interesting experiment in marine propulsion is to be tried soon in this harbor. We allude to the new water jet boat invented by Dr. Walter M. Jackson, of this city. The vessel is 100 feet long, 100 tons burden, with a boiler intended to yield 1,500 horsepower applied to a Worthington pump and used to eject a small stream of water—a three-quarter-inch jet—from the stern post, at the keel line. The water is to issue under the enormous pressure of 2,500 pounds to the square inch, and a speed of between thirty and forty miles an hour is expected by the owners—a velocity far in excess of any other craft afloat."

"At the Eiffel tower, an experiment was performed recently which produced a strong impression on those present. The engineer of the American firm of Otis subjected the Otis lift to a final test before handing it over for public use. The lift was fastened with ordinary ropes, and this done it was detached from the cables of steel wire with which it is worked. What was to be done was to allow the lift to fall, so as to ascertain whether, if the steel cables were to give way, the brakes would work properly and support the lift. Two carpenters armed with great hatchets ascended to the lift; at a given signal, a blow cut the rope and the enormous machine began to fall. Every one was startled, but in its downward course the lift began to move more slowly, it swayed for a moment from left to right, stuck on the brake, and stopped. There was a general cheering. Not a pane of glass in the lift had been broken or cracked, and the car stopped without shock at a height of ten meters above the ground."

"The globe at the Paris Exhibition of 1889 will be forty meters in circumference, and one kilometer will be represented by one millimeter. At this scale, we shall see for the first time upon a globe the place really occupied by certain spaces of known dimensions, such as those of the larger cities. Paris, for example, will occupy upon it nearly a centimeter."



Framework of the great terrestrial globe and its building at the Paris Exhibition

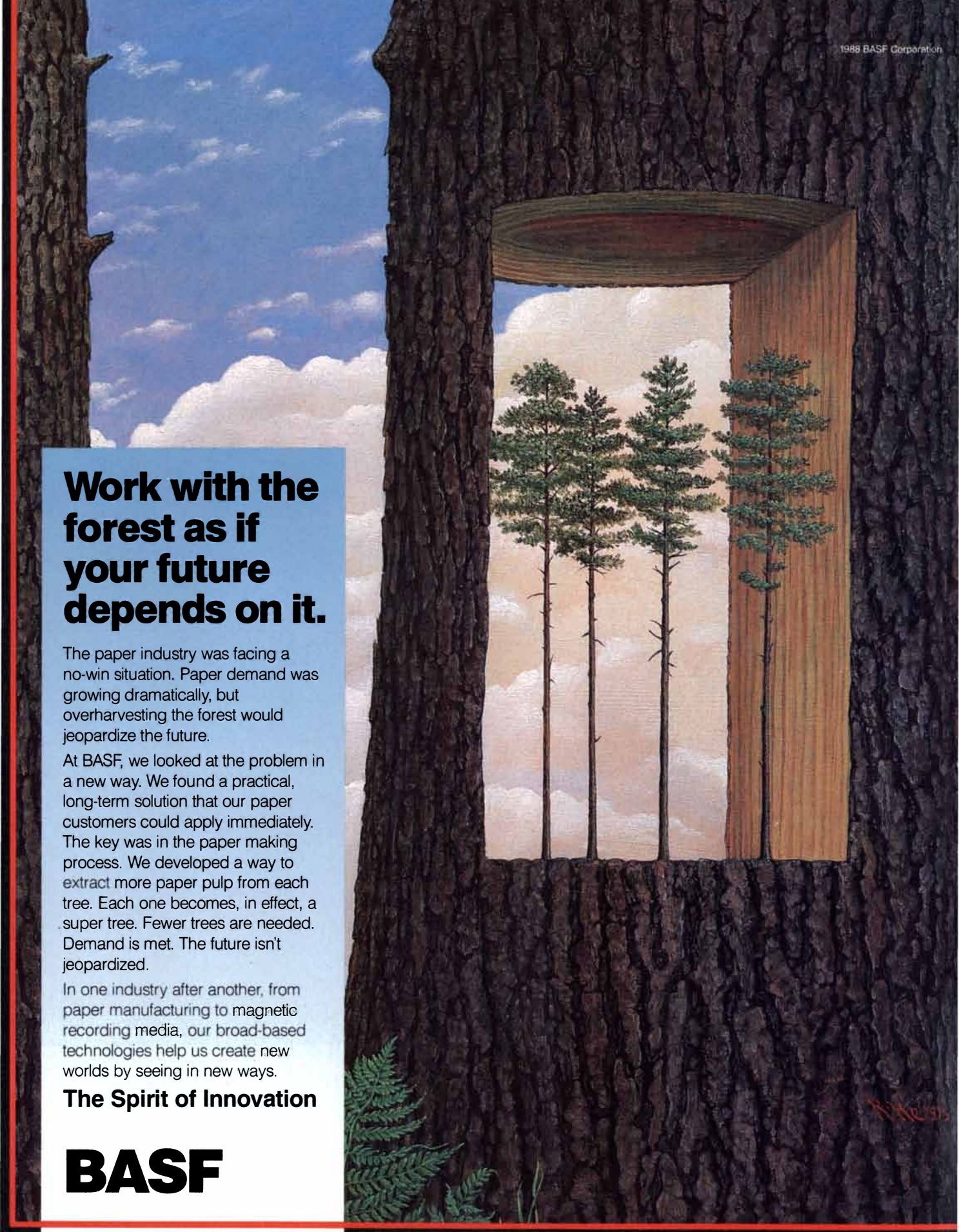


Fact: GM is the only American car-maker to offer a bumper-to-bumper warranty for 3 years or 50,000 miles on every car or light truck we make. See your dealer for details.

GM
MARK OF EXCELLENCE

1989 Buick Regal Gran Sport

Chevrolet, Pontiac, Oldsmobile, Buick, Cadillac, GMC Truck



Work with the forest as if your future depends on it.

The paper industry was facing a no-win situation. Paper demand was growing dramatically, but overharvesting the forest would jeopardize the future.

At BASF, we looked at the problem in a new way. We found a practical, long-term solution that our paper customers could apply immediately. The key was in the paper making process. We developed a way to extract more paper pulp from each tree. Each one becomes, in effect, a super tree. Fewer trees are needed. Demand is met. The future isn't jeopardized.

In one industry after another, from paper manufacturing to magnetic recording media, our broad-based technologies help us create new worlds by seeing in new ways.

The Spirit of Innovation

BASF

Ivan's Eyes

A ban ends, and antisatellite weapons make a comeback

They're back. All but banished by Congressional decree four years ago, antisatellite weapons have returned in force to the wish lists of the Pentagon and to the drawing boards of its contractors. That fact was evident at a recent symposium in Colorado Springs held by the U.S. Space Foundation, a private group that promotes space ventures. Uniformed representatives of all three armed services proclaimed that antisatellite weapons, or ASAT's, are vital to the military's future. Over in the exhibition hall, defense contractors advertised a host of weapons—including lasers and rocket-powered interceptors designed for the Strategic Defense Initiative (SDI)—that could serve as ASAT's. Some exhibiters also offered pithy pro-ASAT apothegms. "It's better to take out Ivan's eyes," said Joe McGowan of the Boeing Aerospace Company, "than it is to take out Ivan."

The Soviet Union, as the Pentagon likes to point out, started it. As early as 1968 the Soviets tested a "coorbital" ASAT: after being lifted into orbit by a rocket, the weapon sidles up to its target by means of small rockets and explodes. In response the U.S. Air Force developed the F-15 ASAT, a heat-detecting missile that is launched by an F-15 jet and then seeks out and rams its target. In 1983, after trying in vain to negotiate a ban on ASAT's with the Reagan Administration, the Soviet Union announced that it was unilaterally halting all ASAT testing. Two years later, over the protests of the White House and the Pentagon, Congress imposed a moratorium on tests in space of the F-15 ASAT. In late 1987 the Air Force canceled funding for the weapon after deciding it was too expensive and its capability too limited. The ASAT issue seemed to be settled.

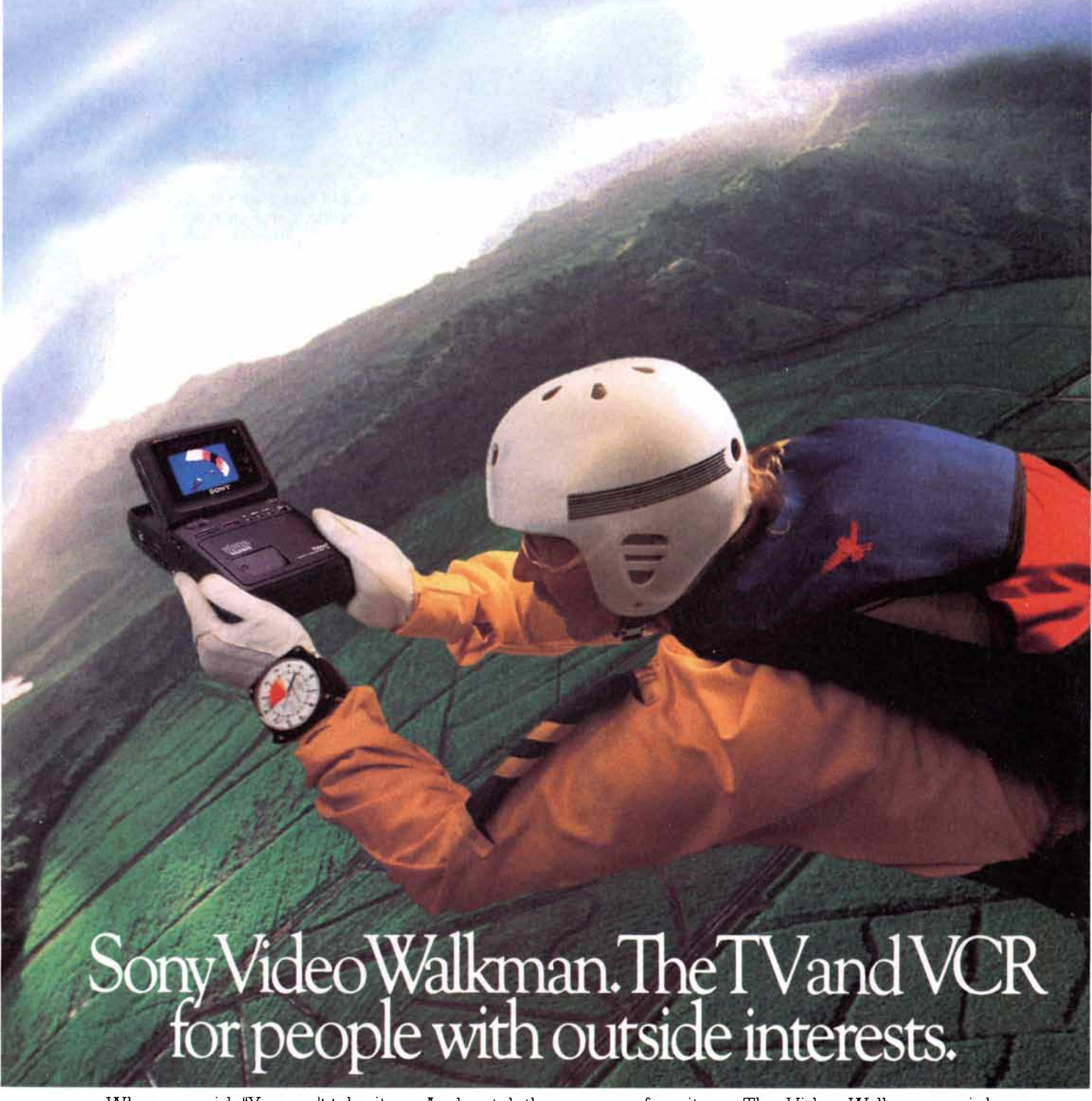
Not quite. Last year members of Congress opposed to ASAT's and to the SDI tried to ban all ASAT tests permanently, and the move backfired: defenders of the weapons mustered enough support not only to reject a permanent ban but also to end the moratorium. Shortly thereafter the Pentagon issued a budget calling for nearly $400 million to be spent on ASAT research and development next year. Each of the services is considering a different SDI system for its ASAT: the Army favors the free-electron laser, the Air Force the excimer laser and the Navy a ground-based interceptor. Yet another SDI-ASAT candidate is the "brilliant pebble": a small, space-based rocket that would supposedly be able on its own to detect, home in on and ram enemy warheads. Developed at the Lawrence Livermore National Laboratory and highly touted by the recently retired SDI director, James A. Abrahamson, brilliant pebbles are an outgrowth of an earlier concept known as smart rocks. (After hearing a briefing on brilliant pebbles at Colorado Springs, Carl Sagan of Cornell University wondered aloud whether "genius dust motes" or "Einstein atoms" might be next.)

Military officials at Colorado Springs maintained that the primary purpose of U.S. ASAT's would be to deter their use by the Soviet Union—or any nation. "The next thing you know it will be some kooky Third-World country that has them," said Rear Admiral Jerry C. Breast, director of operations for the U.S. Space Command. Remarks by other officials made it clear, though, that the armed services (and particularly the Navy, which has long worried about the exposure of its ships to overhead reconnaissance) are also considering the first-use of ASAT's in a wide range of situations. "How low-altitude satellites can be negated by tactical commanders," said Michael I. Yarymovych of the Rockwell International Strategic Defense Center, "is a very serious question that is just beginning to be discussed."

This trend worries such analysts as John E. Pike of the Federation of American Scientists. He points out, as others have for many years, that intelligence satellites have a stabilizing influence on international relations, since they discourage sneak attacks; ASAT's are accordingly destabilizing. The Soviets' coorbital ASAT, which would take hours to destroy a given satellite, does not represent the kind of "bolt-out-of-the-blue" threat that ASAT's under consideration by the U.S. do, according to Pike. If the U.S. develops and deploys these first-strike weapons, he says, the Soviet Union is sure to do the same.

The U.S. has "much more to lose" than the Soviets do by provoking an arms race in ASAT's, according to Senator Tom Harkin of Iowa. In a recent article in *Arms Control Today,* Harkin notes that U.S. satellites are more sophisticated and expensive than Soviet ones; the Soviets' advantage in launching resources would also enable them to replace satellites more quickly. Harkin argues that U.S. security would be

BRILLIANT PEBBLE, a space-based rocket designed for antimissile defense, could also destroy satellites.



Whoever said, "You can't take it with you" didn't have the Sony Video Walkman™ Personal TV/VCR in mind.

It's the first compact portable TV and VCR that can actually be held in the palm of your hand. So if you're someone who's going places, it'll come in very handy.

If you're someone without a lot of time on your hands, there's a built-in 8mm VCR. With an easy-to-use timer. So you can record up to 4 hours of your favorite TV programs. And watch them or your favorite movies in your own good time.

All of which makes the Sony Video Walkman perfect for time-shifting. And place-shifting.

Not only can you use the Video Walkman whenever or wherever you want, you can use it however you want. You can watch its 3" color LCD TV screen with unbelievable sharpness and color fidelity. Or connect it to any other ordinary TV.

But you ain't seen nothin' yet.

The Video Walkman weighs a lightweight 2½ lbs., but it's not light on features. It has high-speed picture search. A sleep timer that turns it off one hour after you've turned yourself off. A worldwide, multi-voltage AC adapter battery charger. And a built-in TV tuner with full VHF and UHF reception.

Which brings us to another feature sure to get a great reception. Our highly adaptable audiovisual input. You'll be sky-high on it because it lets you record and

play back your Video 8® home movies. Or edit them, no matter what format you shot them on.

Now, if you think all this sounds pretty good wait until you hear the Video Walkman. With its exceptional high-fidelity sound.

Besides getting a lot of enjoyment out of the Video Walkman, you'll get a lot of enjoyment into it. For there are over 1300 movie titles now available on 8mm. Like Crocodile Dundee II. The Accidental Tourist. Coming to America. The Dead Pool. And A Cry In The Dark.

And if all that wasn't enough, the sky's the limit on the home instruction, travel and video magazines available. They're part of the video publishing software on 8mm.

So, if your interests take you out. Or in. Up. Or down. Check out the Sony Video Walkman Personal TV/VCR.

It's the most incredible TV and VCR you'll find anywhere.

SONY®
THE ONE AND ONLY.®

far better served by a bilateral ban on ASAT's than by an aggressive pursuit of the technology.

Military officials reject that course. "I don't favor any piece of paper as protection for our assets," Breast says. He contends that the Soviets never stopped testing their coorbital ASAT, despite their proclaimed moratorium, and that they have been developing other types of ASAT's as well, including directed-energy weapons and devices that electronically jam satellites.

The most immediate threat to the military's ASAT plans may, ironically, come from within. For some time the services have squabbled over who should control the ASAT program. In April the Navy, apparently upset by a Pentagon decision that favored the Army, withdrew its request for ASAT funds. Breast says the dispute is only temporary, however, and that shortly the Navy "will be more than eager to move ahead." —*John Horgan*

Taxing the Wages of Sin

Do levies on risky behavior cover the costs to society?

Federal and state taxes on products such as tobacco and alcohol are designed, at least in part, to compensate for the damage done by smoking and drinking. Do they? Analyses by a team of public health researchers from the Rand Corporation, the University of Michigan and Harvard University indicate that smokers probably pay enough in "sin" taxes to make up for their costs to society but that drinkers do not. Taxes on beer, wine and spirits should be raised by about 25 cents a drink (to an average of 50 cents) to bring revenues in line with social costs, according to Willard G. Manning of Michigan.

The result was unexpected. "This started out as a smoking study," Manning said. The group studied the effects of drinking as well to verify their presumption that taxes on cigarettes do not cover the social costs. When the analysis was complete, however, they found that cigarettes are taxed at economically reasonable rates; alcohol taxes, which have not changed substantially in more than 35 years, are "way out of whack," he says.

The medical costs to society of the two habits are comparable, the study indicates, but heavy drinking adds almost a dollar per ounce of alcohol in property damage, court costs and deaths of nondrinkers. (Heavy drinking, defined by the researchers as the amount consumed beyond five drinks a day, accounts, they say, for 40 percent of U.S. consumption.)

Alcohol presents a somewhat more costly demographic picture than does tobacco, according to data gathered at Rand. Although both smokers and heavy drinkers tend to die young, heavy drinkers tend to retire early and so draw as much in pensions, for example, as light drinkers who retire later and live longer. Some financial models, on the other hand, yield a positive societal value for a pack of cigarettes: smokers take more sick leave than nonsmokers and inflate group life-insurance premiums, but they tend to die before costing society much in pensions or nursing-home care. Only when real interest rates rise above about 3.5 percent does the value of those future savings drop enough so that the immediate costs of smoking prevail.

Both alcohol and tobacco consumption correlate with other life choices associated with above-average mortality: these include dropping out of school and not wearing seat belts. If these factors had not been taken into account, they would have skewed the study's results. "Everything we could measure, we controlled for," said Emmett B. Keeler of Rand.

Because drunk driving accounts for so much of the costs of excess alcohol consumption, Manning says, it might be appropriate to tax drinks sold by the glass particularly heavily. For teenagers, he notes, beer consumption decreases with price, so that raising taxes might reduce the number of lives lost as well as compensating for societal costs. —*Paul Wallich*

Playing with Fire

Promethean achievement or scientific opera buffa?

Nuclear fusion at room temperature? If it has been produced—and if it could lead to a practical energy source—the achievement is millennial in its importance. If fusion has not been achieved, the work may take its place as a colossal gaffe, comparable to polywater—a polymeric form of water that was reported in the 1960's and was later debunked. Whatever the upshot of the drama, two electrochemists, Martin Fleischmann of the University of Southampton and B. Stanley Pons of the University of Utah, will have had a profound effect on public perception of science and scientists.

On March 23, Fleischmann and Pons announced to the press, and subsequently their peers through a paper in the *Journal of Electroanalytical Chemistry,* that running an electric current through heavy water using a cathode made of the metal palladium produced an unexpectedly large amount of heat. They said the heat was far in excess of what was supplied as electricity or could be ascribed to a chemical reaction. The workers also detected small numbers of neutrons. The explanation they offered was nuclear fusion.

The workers reported that during the electrolysis the nuclei of deuterium—the hydrogen in heavy water, consisting of a proton and a neutron instead of a single proton—diffuse into the crystalline lattice of the palladium cathode. Fleischmann and Pons claimed that there, packed in great quantity and at high energy, the nuclei overcome their electrical repulsion and fuse, forming tritium (one proton and two neutrons), emitting neutrons and generating heat. Pons told a special forum of the American Chemical Society in Dallas on April 12 that one electrode produced 50 megajoules of excess heat during the course of 800 hours. "There is no conceivable chemical reaction" that could explain anything approaching that amount of energy, Pons asserted: "It's absolutely sure in my mind what we're seeing."

Yet the claims stirred immediate skepticism. Simple calculations suggest that the probability of fusion between deuterium nuclei should be vanishingly small under normal conditions. Moreover, publication by press conference arouses suspicion. Many workers were further dismayed when Fleischmann and Pons withdrew a paper they had submitted to *Nature,* saying they had insufficient time to produce data requested by the journal.

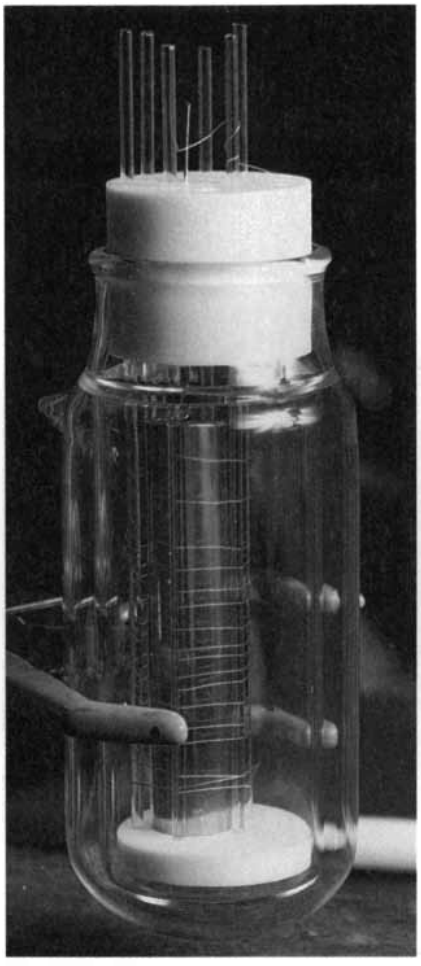
The shadow of Piltdown man has not fallen on the controversy. Nobody believes Fleischmann and Pons are hoaxers, but many workers think Fleischmann and Pons may have made a mistake. The critics point out that the published paper (peer-reviewed by two referees in just a few days) gives few details about the technique or the analysis, and they are especially doubtful about the neutron measurements. Researchers are also irritated by Fleischmann and Pons's failure to do important control experiments before making their announcement. Analyses of the cathodes in search of helium, an expected fusion product, were done only afterward, as were ex-

WE FLY TO
ABERDEEN, ABIDJAN,
ABU DHABI, ACCRA,
ADELAIDE, AMMAN,
AMSTERDAM,
ANCHORAGE,
ANKARA, ANTIGUA,
ATHENS, ATLANTA
AND AUCKLAND.

AND THAT'S ONLY
THE BEGINNING.

A is for an astonishing array of destinations abroad. The most attentive service available on any airline. And an abundance of additional amenities—at the airport and aloft. British Airways. All the best from Aberdeen to Zurich.

BRITISH AIRWAYS
The world's favourite airline.®



FUSION EXPERIMENT done by Martin Fleischmann and B. Stanley Pons consists of a palladium cathode ringed by platinum and suspended in heavy water. Photograph by Terry D. Newfarmer.

periments using ordinary water (which could not sustain fusion).

Critics note many puzzling aspects to the claim. The number of neutrons Fleischmann and Pons report is a billion times smaller than would accompany the reported amount of heat if normal fusion reactions were taking place. Fleischmann and Pons suggest that an exotic neutron-free reaction producing helium-4 (which they have also detected) takes place instead, and theorists have obligingly proposed ways exotic fusion reactions might be favored. Other observers suspect less exotic phenomena. Richard L. Garwin of the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y., in a critique published in *Nature,* suggests that imperfect stirring or localized chemical reactions in the vessels of heavy water could easily have deceived the workers.

At industrial, academic and government laboratories around the world, scientists scrambled to reproduce the experiment, communicating by FAX, telephone and computer. In spite of the frenzy, by late April only a handful of groups had reported any evidence of success. Workers at the Lajos Kossuth University in Hungary and at Moscow M. V. Lomonosov State University reported detecting neutrons but did not measure heat. (Workers at the Georgia Institute of Technology withdrew a claim of neutron detection when they discovered their detector was actually reacting to temperature changes.) John A. Wethington, Jr., and Glen J. Schoessow of the University of Florida claimed to have found clear evidence of the fusion product tritium in their experiment.

Charles R. Martin and his colleagues at Texas A&M University instead claimed sustained excess heat production: "There is no doubt that we are seeing excess energy," Martin says. Robert A. Huggins and his colleagues at Stanford University announced in April that whereas heavy water generates excess heat, ordinary water does not. "The effect is real and it is substantial," he said, adding that the excess heat could therefore not be from a chemical reaction.

Michel W. Barsoum and Roger D. Doherty at Drexel University, on the other hand, said they found unexpected heat production with both ordinary and heavy water. Pons says that some 60 other laboratories have obtained partial confirmation but want to keep the news quiet for now. (The University of Utah has made extensive patent claims.) Fleischmann explains that he and Pons were pressured into making their announcement because word was leaking out. Pons suggests that unknown metallurgical factors may explain other workers' low rate of success.

Fleischmann and Pons had originally planned to publish their submission to *Nature* alongside one by rivals Steven E. Jones and his colleagues at Brigham Young University and Johann Rafelski of the University of Arizona. That paper, which also describes cold nuclear fusion in metals, did appear in *Nature* on April 27. Like Pons and Fleischmann, Jones and his colleagues experimented with metals that absorb deuterium during the electrolysis of heavy water. The workers detected neutrons, which they present as evidence of fusion, but the reactions occurred at too low a rate to generate measurable heat.

It is Fleischmann and Pons who have captured public attention, however. Only time will tell whether their claimed effect is real. In *Nature* Garwin writes: "Large heat release from fusion at room temperature would be a multi-dimensional revolution. I bet against its confirmation." If Garwin is right, public regard for science could be singed. —*Tim Beardsley*

Waiting Game

The Gulf of Alaska will recover, but not quickly

How long will it take the Gulf of Alaska to recover from the ecological nightmare that began in the first minutes of March 24, when the *Exxon Valdez,* loaded with 1.2 million barrels of crude oil and headed for the open sea, ran instead onto a reef in Prince William Sound? Can the plant and animal life of the region ever regain its pristine richness?

History suggests there will be an eventual recovery, although it could take a decade or more. The long-term consequences of spills vary widely, depending on the type of coastline affected and on wind and temperature conditions. The effects of the *Amoco Cadiz* spill, which released 1.6 million barrels of oil onto the coast of France in 1978, were still evident eight years later, and sensitive environments such as marshes are only now returning to normal. Yet the giant blowout at *Ixtoc I,* an offshore rig, which released about a million tons of oil into the Gulf of Mexico in 1979, had fewer obvious long-term effects: storms broke up much of the oil, which was released over a period of months, and relatively little reached sensitive coastlines.

The circumstances in Alaska, where 250,000 barrels were spilled, are very different. For one thing, it is cold, and there is little experience with the effects of oil in subarctic conditions, according to John W. Farrington of the University of Massachusetts at Boston, who has studied several oil spills. The surface water in Prince William Sound was a chilly 3 degrees Celsius in mid-April. At that temperature evaporation is slowed, and degradation by microorganisms, which ultimately removes much spilled oil, takes twice as long as it does even at 10 degrees, according to Don K. Button of the University of Alaska at Fairbanks. Some researchers, such as Edward J. Brown, also at the university, have suggested spraying the shoreline with oil-eating bacteria, which are grown in large tanks for treating tanker-ballast water.

For its part, oil has multiple effects on living things. Birds are conspicu-

New primers help protect electronic circuits in neural prosthesis devices. The primers, plasma polymerized hydrocarbon films developed by Hughes Aircraft Company, are used to bond biologically inert protective coatings to the devices for periods of over 10 years. The prosthesis devices help victims of neural trauma, such as stroke, to regain some of their lost neural functions by electrically stimulating proper areas of the brain. Plasma polymerized films are also useful as protective coating in many other applications such as infrared optics that are exposed to extreme conditions of sand and salt water.

A new fiber-optic modem is the first non-cryptographic communications security product to be endorsed by the U.S. National Security Agency. It is approved for the protection of all levels of classified data. The FAM-131 modem is part of an intrusion detection optical communications system (IDOCS), developed by Hughes, that operates on the principle of alarmed fiber optics rather than conventional encryption. With no encryption signal to interfere with the data stream, the IDOCS provides high signal transparency, and can operate up to 13 megabits per second. An additional benefit is the elimination of crypto key management. IDOCS is intended for use between buildings in campus-like environments and local area networks for various applications including command and control and information management.

An advanced concept in helicopter dipping sonar will provide a significant increase in anti-submarine warfare (ASW) capability. Under development by Hughes for the U.S. Navy's Airborne Low Frequency Sonar (ALFS) program, the new sonar is designed to operate from both the LAMPS MKIII and the CV helicopters. This will enhance the Navy's Airborne ASW effectiveness in both inner and middle zones of battle group deployment. Hughes, and its teammate Thompson Sintra ASM, will demonstrate the performance of this new sonar in flight tests aboard an SH-60B helicopter in early 1989. The team's ALFS concept embodies leading-edge technologies in the areas of high gain acoustic arrays, electro-mechanical kinematics, fluid dynamics, signal processing and man-machine interfaces.

The integrated helmet display system for the next generation of multimission helicopters will give pilots a "head-up" and "eyes-out" attitude. The display system, a crucial part of the U.S. Army's Light Helicopter Experimental (LHX) avionics package, will allow pilots an 80-degree field of view and display all primary aircraft functions, such as attitude, altitude, speed, and other attributes. Hughes is developing the display system with Honeywell. Hughes and Texas Instruments will also develop the LHX's target acquisition system, night vision pilotage system, computer system, and communications subsystem.

Hughes Research Laboratories seeks highly-qualified scientists for advanced research in physics, chemistry and electronics. Disciplines include: Information sciences and artificial intelligence; space plasma sources; pulsed power switches; free electron lasers; electron beam testing; advanced IR detectors; liquid-crystal materials and displays; nonlinear optics and phase conjugation; computer architectures for image and signal processors; GaAs microwave devices and IC technology; and optoelectronic materials and devices. Send your resume to: Professional Staffing, Hughes Aircraft Company-Research Laboratories, Dept. S2, 3011 Malibu Canyon Road, Malibu, CA 90265. Equal opportunity employer. Proof of U.S. citizenship required for some positions.

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

HUGHES

Subsidiary of GM Hughes Electronics

ously affected, along with fur seals and sea otters, which rely on their pelt to maintain their body temperature. Mammals such as harbor seals that are insulated by blubber are much less sensitive, says Mark A. Fraker of BP Exploration, Inc. Fish escaped immediate harm, reports David Kennedy of the National Oceanic and Atmospheric Administration, who is organizing scientific assessments in Valdez. That could change. For one thing, fish populations will be reduced if the usual spring bloom of plankton (on which fish feed) is badly affected. Most small organisms that came into contact with the oil were quickly killed.

Kennedy predicts oil will remain visible in sheltered areas for several months. Short-term consequences, Farrington suggests, could result more from aromatic chemicals such as xylene, toluene and benzene than from the oil itself. Long-term effects will result from these and from oxidized aromatic breakdown products, which dissolve in cell membranes, impair growth and are thought to cause genetic damage as well. Alaskan crude oil contains high levels of multiplering molecules, large molecules that take a long time to degrade.

Farrington estimates that ecological disruption will last for from five to 15 years. The organic chemicals, he notes, will be replenished from oil trapped in pockets between rocks and in bottom sediments. Oil has in some areas penetrated several feet into spaces between rocks. There it can persist in its unweathered state for a long time, according to Kennedy. Farrington has some hope that the glacial sediments of the sound may prove less absorbent than some other deposits and so allow the oil to be diluted faster.

Eric Schneider, who is leading the National Oceanic and Atmospheric Administration's assessment, says that in other spills life has returned in stages: within weeks phytoplankton appear, then the zooplankton that feed on them, followed by larger organisms. One of Schneider's chief concerns is that low levels of organics can disrupt the chemosensory organs of fish and invertebrates. Such an effect could mean bad seasons ahead for local fisheries. When fiddler crabs are poisoned with aromatics, for example, they dig their burrows at shallow angles, and so they perish. Moreover, aromatic toxins are concentrated in the food chain, so that when animals do return, their flesh may be tainted.

David G. Shaw of the University of Alaska at Fairbanks observes that the great Alaskan earthquake of 1964 was a bigger disruption of marine life and that recovery was eventually complete. It does take time—time that will pass slowly for Alaskans who make their living from the sea and for all people who treasure nature. —*T.M.B.*

PHYSICAL SCIENCES

Nonlinear Thinking

Pentagon strategists take heed of a dove's chaotic theories

Gottfried Mayer-Kress seems an unlikely person to have won the respect of strategists in the U.S. Department of Defense. In West Germany, where he was born and trained as a physicist specializing in nonlinear dynamics, he voted for the Green Party and demonstrated against the deployment of nuclear missiles by the U.S. Even after taking a job at the Los Alamos National Laboratory in 1984, he continued his antinuclear activism: for several years on the anniversary of the U.S. bombing of Hiroshima, he passed out nuclear-disarmament leaflets to startled colleagues, many of whom design nuclear weapons.

Nevertheless, during the past year or so researchers from the Defense Intelligence Agency, the U.S. Naval Academy and other military institutions have sought Mayer-Kress's counsel. They are interested not in his political views but in his assertion that nation-states can be described mathematically as nonlinear and even "chaotic" systems; in such systems small perturbations—the deployment of antimissile defenses by the U.S., say, or the bombing of a key communications link by terrorists—can trigger large and unpredictable consequences—such as nuclear war. "I see a lot of potential in his ideas," says Major Roy E. Rice, who heads a group that does research on modeling techniques for the Pentagon's Joint Staff.

Mayer-Kress joined Los Alamos (he holds a joint position at the laboratory's Center for Nonlinear Studies and at the Santa Fe Institute, a group with close ties to the center) to study purely scientific issues. Feeling somewhat guilty about working in a weapons laboratory, he was seeking a way to apply his talents to the cause of peace when a paper in *Nature* caught his eye. The author, Alvin M. Saperstein of Wayne State University, suggested that the same mathematics that described the transition of a jet of water from

Slow road to glory

The Story of the Lands' End Rugby Shirt

by Red Mulcahy

We had high hopes for our original Rugby Shirt when we introduced it in 1980.

After all, it was heavyweight cotton. Not a bit flimsy. And darn good-looking too, with its jaunty stripes. (We called 'em Team Stripes, which we thought had a real ring of authenticity.)

So we mailed out our catalogs, and waited for the applause.

Setback Number One.

What we got was more like a sustained Bronx cheer. Our customers complained that our rugby shirts shrank too much. Up to 20%. Went in the dryer a Large, came out a Medium.

We were embarrassed. Took the shirts out of our catalog, and even wondered whether we should leave rugby shirts to the sporting goods companies.

But our feisty nature got the best of us. We went back to the drawing board and developed a preshrunk 100% cotton jersey fabric. A beefy fabric, 10.5 oz. versus the usual 5 to 9 oz. A fabric that reduced shrinkage to a tolerable 3%.

So far, so good. But we wanted to be sure our improved shirts were the real thing. Especially since about this time, lots of "rugby shirts" were appearing on the market that were really nothing more than colorful sportshirts.

We figured a "field test" was in order. And gave our shirts to the University of Chicago rugby team. (They happened to be close at hand.)

Setbacks Number Two and Three.

The results were disastrous. While our new fabric stood up, almost nothing else did. A particular problem was the two-piece placket, which the ruggers consistently tore asunder.

Again, we went back to the drawing board, and developed a more rugged continuous placket—all one piece, with no weak point.

Would this make our shirt tough enough for rugby? We decided to submit our latest shirts to the ultimate test—international rugby—giving them to the USA Eagles, America's national team. The biggest, toughest rugby players in the country. (Some of these guys were born with five o'clock shadow.)

OUR SHIRTS FAILED AGAIN! Seams came apart. Buttons popped off. In a Hong Kong match, one Eagle came off the field wearing nothing above the waist but a collar.

We went (no, we trudged) back to the drawing board. But this time, we had the help of the Eagles. We switched to stronger thread. Developed deeper biting seams. Beefed up our shirt at collar, placket. And added other indomitable (we hoped) features.

The Eagles took our shirts into action in June 1984 against Canada. And we held our breath.

At long last: "Tough as the game!"

To our relief, these Lands' End Rugby Shirts finally stood up to all the bruising, tugging, grabbing, tackling punishment the Eagles could put them through, in some of the toughest international competition imaginable.

And have continued to stand up, year after year. In 1987, they even survived three tough matches at the Rugby World Cup in Australia.

Now, that makes us pretty proud. Especially considering where these shirts came from.

Just shows what hard work (and a little humiliation) can do.

A Postscript...

If you'd like to put our "tough as the game" Rugby Shirt (or any of our other clothing and soft luggage) to your own "field test," write or call today for a free Lands' End catalog.

You'll discover more than a great shirt—you'll discover a new direct way of shopping in which Quality, Value and Service are absolutely, positively GUARANTEED. PERIOD.®

laminar to turbulent flow might be employed to describe the outbreak of war between nations.

Intrigued, Mayer-Kress contacted Saperstein, and over the next two years they developed a model that simulated how the deployment of a space-based antimissile defense—such as the Strategic Defense Initiative envisioned by Ronald Reagan—might affect relations between the U.S. and the Soviet Union. The model incorporated nonlinear equations indicating how actions by one side—an increase in spending on antimissile satellites, for example—might spur a response by the other side that in turn provokes a counterresponse. To simulate the "a priori unpredictability of human decisions," Mayer-Kress and Saperstein included a function that causes some variables to fluctuate at random. Then they ran the model on a supercomputer and found that an SDI-type system would probably trigger an arms race in both offensive missiles and antimissile defenses.

They summarized their conclusions in the Center for Nonlinear Studies' newsletter in 1986. Predictably some Los Alamos officials were upset, but they did not censor Mayer-Kress. Last December the *Journal of Conflict Resolution* published his and Saperstein's paper, "A Nonlinear Dynamical Model of the Impact of SDI on the Arms Race." Two months later *Nature* published a paper by Mayer-Kress and a colleague from West Germany: "Chaos in the International Arms Race."

Even before these papers were published, military analysts outside Los Alamos had begun to contact Mayer-Kress. One visitor, an official in the Defense Intelligence Agency who requested anonymity, told SCIENTIFIC AMERICAN that Mayer-Kress's nonlinear method could help the Pentagon uncover vulnerabilities in its command and control network—and in the Soviets'. The agency may also use the method in classified studies of the impact of AIDS on the stability of Third-World governments and of the effect of military intervention on drug trafficking, according to the official.

Robert Artigiani of the U.S. Naval Academy, who persuaded Mayer-Kress to address a meeting at Annapolis in March, says chaos theory might help the Navy determine whether its forces are too vulnerable to technological change—represented, for example, by the advent of a cheap, long-range antiship missile. Indeed, Mayer-Kress's ideas mesh with a growing trend in the Pentagon toward modeling not only traditional war-fighting scenarios but also "nonfirepower-driven" issues, according to Major Rodney B. Mitchell of the Defense Communications Agency. "We know what happens when the bomb hits," says Mitchell, who has borrowed Mayer-Kress's model to study the effects of Soviet conventional-arms reductions on the balance of power in Europe. "Now we want to know what happens before it hits, or if there are no bombs."

That, says Mayer-Kress, is exactly the point. "We all want the same thing," he says, "to avoid war." —*J.H.*

Muddy Evidence

It keeps alive the case for life on Mars

The hope (however dim) of discovering life on Mars may lie buried in mud, or suspended in water, somewhere beneath the planet's surface. Proponents of the possibility that life may be possible on Mars have had to contend with the fact that liquid water would at best enjoy only a fleeting existence in the extreme cold of our outer solar neighbor. Typical winter temperatures on Mars are about −123 degrees Celsius—cold enough to freeze carbon dioxide in the polar regions. Vitalists have therefore pinned their hopes on images captured by *Mariner 7* in 1971 and by subsequent spacecraft that seem to show "river-made" erosion.

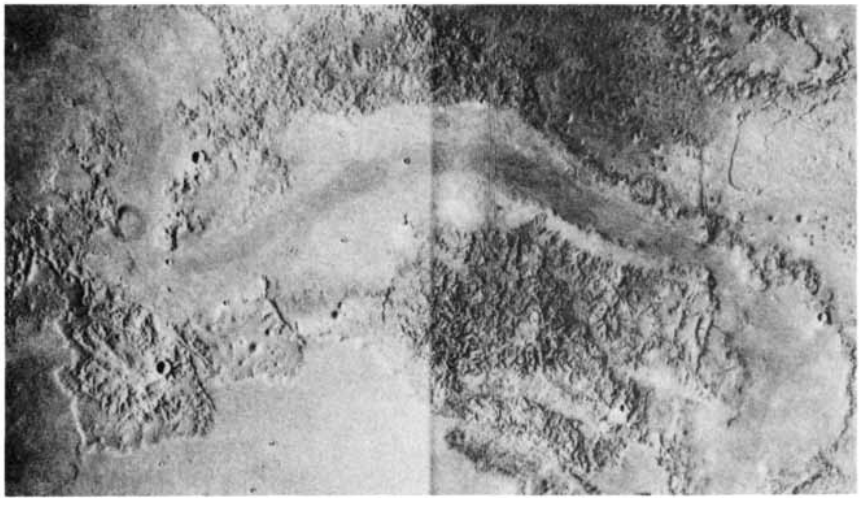
Reanalysis of Viking orbiter images of the western Elysium volcanic province of Mars, recently reported in *Geology* by Eric H. Christiansen of Brigham Young University, strengthens the case for Mars as a water-rich planet and therefore as a possible nursery for life. Christiansen's interpretation of geomorphic evidence—channeled flows of volcanic debris recorded in the satellite photographs—reinforces the theory that water is, in fact, present below the surface of Mars.

Collectively called lahars (the Indonesian word for volcanic-debris flow), these channeled deposits in Mars' volcanic region seem to be gravity-driven mass flows, similar to those seen at Mount St. Helens in Washington: they run downhill, led by a snub-nosed front that leaves coarse-grained lateral sediment in its wake.

Lahars are wet, close mixes of water and solids. Their pattern of troughs, channels and pitting points to subsurface water. Photographic evidence of the Martian topography suggests that just as the Washington volcano's lahars were 22 to 36 percent water by volume so too were the now dry, channeled deposits present on the Elysium volcanic dome. Branching implies that seepage valleys were cut as water was expelled to the surface; drainage also seems to have generated certain shapes, whereas other shapes suggest evaporation. Christiansen reports that the Martian channels and deposits extend from a system of fractures, which also fed lava flows on a vast scale.

In his view the debris flows indi-

Satellite photographs of volcanic flows suggest that water underlies the surface of Mars



COMPOSITE OF TWO PHOTOGRAPHS of an area 90 kilometers across, made during Viking orbiter missions, shows deposits left on Mars by volcanic flows.

INDULGENT. THE SENSE OF REMY.

Exclusively Fine Champagne Cognac

Rémy

cate massive volcanic action, which generated mud flows, depositing sediment in patterns similar to those found on the earth. Christiansen argues that volcanism and the lahars—with attendant water—were close in space and time.

Might these surfaces simply be the reworked configurations of dry, volcanic avalanche deposits? Christiansen answers that the ratio of "source scar" to deposit volume is equal in such deposits on the earth; the Martian lahars have from 10 to 100 times the volume of the troughs from which they appear to have emanated, further indicating that both debris and water issued from below the surface.

Christiansen's present working hypothesis is that subsurface ice was melted by regional volcanic action. He considers the frozen mud flows to be further evidence that subsurface water was once at work (and perhaps still is present) on Mars. The lahars—never identified as such before—are explainable by nothing less than volcanic temperatures acting on extensive buried reservoirs, generating both water flows from the subsurface ice and steam from within the planet's crust. *—Barth David Schwartz*

Halley's Birthplace

The comet may be a drifter from the fringes of the galaxy

Was Halley's comet born in the cradle of the primordial solar system, or is it an interloper of uncertain pedigree from the outer reaches of the galaxy? According to prevailing theory, water, carbon monoxide and other matter condensed, coalesced and froze to form Halley's comet in the rotating disk of interstellar gas and dust from which the solar system also emerged about 4.6 billion years ago. The comet then settled into an orbit between Neptune and Uranus. In the course of billions of years the gravitational forces of passing stars and the planets shifted the comet into a 76-year orbit that passes as close to the sun as Venus.

Workers at Arizona State University have produced a piece of evidence that discomfits the local-origin theory. When Halley's comet hurtled by the earth in 1986, Susan Wyckoff and her colleagues at Arizona State observed the comet's coma, or atmosphere, through a telescope on Mount Stromlo in Australia. By analyzing the emitted light with a spectrometer, the group determined the relative abundance of two carbon isotopes in a carbon compound that was blowing off the comet's surface.

Wyckoff reports in the *Astrophysical Journal* that the comet contains about one part of carbon 13 to 65 parts of carbon 12. This proportion is about the same as what is measured in interstellar gases in the periphery of the galaxy. Samples from the earth and moon, meteorites and the atmospheres of the outer planets contain about one part of carbon 13 to 89 parts of carbon 12.

The resolution of the question of Halley's origin depends, Wyckoff says, on a clearer understanding of how or whether the ratio between carbon isotopes has changed in the galaxy during the time that has elapsed since the solar system's birth. During that period carbon 13 spewing from red giant stars may have enriched the interstellar medium of the outer galaxy while the proportion in the solar system remained constant. In that case, Halley's comet might have coalesced in the outer galaxy some time after the birth of the solar system. The gravitational force of the sun and planets then captured the comet. Capture is a remote possibility but not one that can be ruled out, Wyckoff says.

Supporters of the local-origin hypothesis can shore up their position by presuming that carbon 12 from a stellar explosion bombarded a part of the primordial solar system that formed the bulk of the modern solar system; Halley's comet, however, emerged from a region poorer in carbon 12. This hypothesis, workers say, is also improbable. Therefore, the issue may have to rest until the comet's next visit, expected in 2061. *—Russell Ruthen*

BIOLOGICAL SCIENCES

Disco-Bee

Scientists build a "bee" whose dancing fools real bees

It beats its wings 250 times a second. It waggles its rear end 15 times a second. It skitters in tight figure eights on the inner wall of a hive. It even spews sugar water from its mouth. But it is not a real dancing bee—it is a robot built by scientists in Europe. It also a "real breakthrough" in the study of bees, according to James L. Gould of Princeton University.

The German biologist Karl von Frisch discovered more than 30 years

ago that bees employ a "waggle dance" to inform one another of the location of nectar-bearing fields. The orientation of a dancer's figure eights indicates direction, and the duration of its dance indicates distance. Since then many investigators, including Gould, have tried in vain to build a fake bee that could dance so well that real bees would follow its directions. The investigators hoped such a device might help them determine more precisely how bees communicate through choreography.

After three years of effort, Wolfgang H. Kirchner of the University of Würzburg in West Germany and Axel Michelsen of Odense University in Denmark finally succeeded where others had failed. They forged a body of bronze and cut wings from razor blades. The wings pivot on diamond bearings. The robot is not freestanding: it is linked by wires and a rigid tube protruding from its back to computer-controlled servomotors and a reservoir of sugar water.

The scientists put the robot through its paces last summer at a field station in Würzburg. After being inserted into a hive, the robot danced out directions to a specific region and served up samples of sugar water. Sure enough, some of its audience flew off to the designated area. The robot did not "recruit" as many bees as a real dancer usually does, but Kirchner and Michelsen hope a second-generation robot will perform even more convincingly.

The robot has already provided an important scientific payoff. Its inability to recruit any bees when its wings were not beating confirmed what many researchers had suspected: bees follow a dancer's movements in the dark hive by "listening"—probably by means of specialized organs in their antennas—to changes in the amplitude of the dancer's buzzing. At close range the buzzing is comparable in intensity, Kirchner notes, to "a very loud discotheque." —*J.H.*

Grave Doubts

The Neanderthals may not have buried their dead after all

Did the Neanderthals bury their dead? Since the beginning of this century the widely accepted answer has been that they did. Now Robert H. Gargett, a graduate student at the University of California at Berkeley, has written an article arguing that all the known Neanderthal "burials" can be accounted for by natural preservational processes. If the Neanderthals did not bury their dead, then their spiritual capacities may have been overrated; perhaps they were not much like modern human beings. Gargett's article, published in *Current Anthropology,* has evoked a broad range of responses from his colleagues.

At issue is the question of what enables an anthropologist to conclude that a particular skeleton was buried. There is little doubt that beginning about 60,000 years ago a number of relatively complete Neanderthal skeletons begin to appear in the archaeological record of Europe—in sharp contrast to preceding periods, when the known skeletal remains are fragmentary. The discovery of skeletons at sites such as La Chapelle-aux-Saints and La Ferrassie in France led early anthropologists to conclude that the Neanderthals had consciously interred their dead.

Gargett, however, proposes that the differential preservation of Neanderthal remains may be due to geology and not human agency. All the more or less complete skeletons, he points out, were found in caves, where complex geologic conditions might mimic interment. At La Chapelle-aux-Saints, for example, where a nearly complete skeleton was found in 1908, the roughly rectangular depression in which the bones lay might have resulted from the dissolving action of water on the limestone floor of the cave. Had a weakened Neanderthal crawled into the small cave and then died, the natural deposition of sediments would have left something resembling a grave.

At La Ferrassie, a rock-shelter near the Dordogne River where two almost complete and five fragmentary skeletons were found between 1909 and 1920, natural forces might also have imitated burial. The feature that has stimulated the most discussion about La Ferrassie is a group of nine mounds, one containing remains of a Neanderthal. The original investigators concluded that the mounds were built in an interment ritual, but Gargett suggests they may have been due to the action of frost, which can create geometrically patterned hummocks. "That the eight other mounds were 'empty' suggests some explanation other than human agency," he writes.

The comments on Gargett's article, published in the same issue of *Current Anthropology,* reveal a startling spectrum of responses from specialists. Geoffrey Clark and John Lindly of Arizona State University say Gargett

"successfully dismantles the argument for intentional burial of the dead at several well-known Eurasian Neanderthal sites." David W. Frayer and Anta Montet-White of the University of Kansas, however, say Gargett is "selective in the cases he covers, lacks historical perspective and ignores important discoveries that demolish his argument."

Gargett said recently that his aim had not been to dismiss the possibility of Neanderthal burial altogether. Instead, his intention was to look skeptically at the Neanderthal "burials" in order to better understand the emergence of modern human culture. Some investigators think that the Neanderthals were our ancestors and that the transition to modern culture resulted from the evolution of Neanderthals into anatomically modern humans. Much recent evidence, however, suggests Neanderthals may not have been our ancestors and that the appearance of modern humans preceded both the Neanderthals' disappearance and the emergence of anything resembling modern cultural patterns.

"I was addressing preconceptions about Neanderthals' cognitive abilities," Gargett said. "After all, their brains were as large as ours. Their tools were like those of anatomically modern humans living elsewhere in the world at the same time. If they also buried their dead, they begin to seem much like us, and it becomes difficult to understand the differences between them and us. If you entertain the possibility that they did not bury their dead, however, their spiritual capacities may not have been so much like ours. That would make it easier to understand some other things—such as why they apparently produced no art." —*John Benditt*

The Tails of Ubiquitin

The molecule is linked to both protein decay and synthesis

The small protein ubiquitin is a versatile actor. It pervades all nucleated cells, binding to other proteins in chromosomes, the cellular skeleton and surface receptors, and its roles are as diverse as the company it keeps. Furthermore, the protein's structure has been conserved to an extraordinary degree in the course of evolution; yeast ubiquitin differs from human ubiquitin by a mere three amino acids.

To some molecular biologists, this remarkable conservation suggests that virtually every part of the 76-amino-acid molecule must be involved in processes vital to all nucleated cells. It is known, for example, that ubiquitin plays an important role in the destruction of defective proteins and of regulatory proteins that have served their purpose. Now it turns out that ubiquitin also promotes the function of certain proteins making up the ribosomes, the cellular factories that assemble proteins according to the blueprint provided by messenger RNA.

The discovery, by Daniel Finley, Bonnie Bartel and Alexander Varshavsky of the Massachusetts Institute of Technology, arose unexpectedly in the course of their investigation into ubiquitin's role in protein destruction. Chains of ubiquitin had been found to bind to proteins destined for elimination, where they apparently signal a mechanism to break down the proteins. The investigators identified four sites in yeast DNA that code for ubiquitin. Three of the sites, called *UBI1, UBI2* and *UBI3,* each consist of a single ubiquitin gene fused to a "tail" gene encoding a second protein. The *UBI1* and *UBI2* tails both proved to encode a specific ribosomal protein with 52 amino acids. The *UBI3* tail encodes a ribosomal protein consisting of 76 amino acids. (The fourth site, *UBI4,*

codes for a chain of six ubiquitins.)

The ribosomal proteins encoded by the tail portions are essential for cell survival. When the *UBI3* gene was deleted from yeast DNA, the growth rate of the cells plummeted; deleting both *UBI1* and *UBI2* killed the cells outright. When the investigators snipped off the ubiquitin genes proper and reinserted just the tails into yeast cells, the cells revived. On the other hand, the ubiquitin portion of these genes seemed nonfunctional: when the workers reinserted only the ubiquitin portion, the cells were impaired or killed anyway.

Why, then, are the tails fused to the ubiquitin gene? When the investigators reinserted the tail portions alone, they found that multiple copies were needed in order to achieve the same level of function normally ordained by a single fused ubiquitin-tail gene. They surmise that although the fusion of ubiquitin to ribosomal proteins is not vital to cell survival it does enhance the proteins' function. The fusion arrangement is conserved through evolution with as much fidelity as the ubiquitin molecule itself. Kent L. Redman and Martin Rechsteiner of the University of Utah School of Medicine recently reported that human ubiquitin genes are fused to ribosomal-protein genes nearly identical to those in yeast.

Perhaps the ubiquitin stabilizes the ribosomal proteins or assists in their incorporation into the ribosome. The M.I.T. researchers note that ubiquitin is very small and stable; they think its temporary attachment to certain proteins may serve to stabilize them during a critical juncture in their biological activity. It may be that ubiquitin first evolved to carry out some essential role—perhaps its established function as a tag for protein destruction—along with enzymes that cleave off the ubiquitin after it has done its job. The ubiquitin and its associated enzymes may then have evolved a diversity of new functions involving other proteins. —*June Kinoshita*

Molecular Monkeywrench

Synthetic nucleic acids can block gene expression

Sophisticated techniques for synthesizing genetic molecules are suggesting new ways to destroy tumor cells or cells infected with a virus. Synthetic nucleic acids can interfere either with a cell's DNA or with its messenger RNA's, the molecules that carry the genetic message from the genes to where proteins are made. Because the genetic instructions in an infected or malignant cell differ from those in a healthy cell, the jamming technique can be designed to spare healthy cells.

The concept is elegantly simple; realizing it has been difficult. Several years ago investigators tried putting into cells a synthetic single-strand DNA comprising a sequence of chemical bases that matched the messenger RNA (mRNA) of a particular functioning gene. The synthetic DNA bound to the mRNA. With the mRNA thus inactivated, the cell was prevented from making the protein encoded by the gene. The synthetic nucleic acid is called "antisense," because it has a base sequence complementary to the sense-making natural gene.

Yet there were difficulties: cellular enzymes quickly destroyed the antisense nucleic acid, and cells had to be exposed to high concentrations of the synthetic DNA before they would take up enough to have an effect.

More recently advances in chemical synthesis have allowed workers to make modified RNA's and DNA's that resist attack by cellular enzymes and pass more easily through cell membranes. One such modification con-

sists of adding methyl groups (CH_3) to the antisense DNA along its phosphate backbone. Paul Miller, a pioneer in the field, Paul Ts'o and their colleagues at the Johns Hopkins Medical Institutions are investigating such compounds, which they call matagens (for MAsking TApe for Gene ExpressioN). They and Laure Aurelian of the University of Maryland found that a matagen directed at a herpesvirus gene can block the virus both in human cell cultures and in the skin of mice.

Antisense DNA can also be modified so that sulfur atoms stud its phosphate backbone. Makoto Matsukura and Samuel Broder of the National Cancer Institute have made such "phosphorothioate" antisense compounds that are targeted at the *rev* gene of the AIDS virus. The phosphorothioate inhibits expression of the virus in chronically infected cell cultures in the laboratory. Moreover—mysteriously—it prevents infection of new cells, according to Matsukura. Their work is published in the *Proceedings of the National Academy of Sciences*. Matsukura cautions that trials to determine effectiveness in animals are still some way off. Even if the trials are successful, he says, phosphorothioates may fail as a drug: patients could develop antibodies to the compounds, which could precipitate an autoimmune disease.

A variation on the theme employs synthetic nucleic acids that bind not to the mRNA but to the double-strand DNA itself, forming a three-strand DNA. The extra strand keeps the gene from making mRNA in the first place. This approach might be used to fight latent viral infections or to inhibit oncogenes.

In the *Journal of the American Chemical Society* Thomas J. Povsic and Peter B. Dervan of the California Institute of Technology report that they can now combine chemically altered bases to make nucleic acids that bind to specific sequences of double-strand DNA under the conditions that prevail in cells. The synthetic third strand binds well enough to obstruct proteins that normally regulate the activity of the DNA. According to Dervan, the technique might be applied for mapping DNA and isolating genes as well as for genetic targeting.

Dervan and others have developed synthetic chemical structures that cut DNA once they are bound to it—which would inactivate any gene that encompassed the break. The possibilities of such work have not been lost on the growing number of pharmaceutical companies that are now pursuing various avenues of antisense research. —*T.M.B.*

OVERVIEW

One Hand Clapping

Digital audio is changing film sound. Will audiences notice?

Visually, the shot of a meat-packing warehouse and its gritty environs in the Bronx is spare and detached. It is the sound that gives the scene its oppressive immediacy: the tandem thumps of traffic crossing a metal joint in the elevated highway overhead. Like so many "familiar" sounds in movies, the effect in Sidney Lumet's forthcoming movie *Family Business* is in fact a composite of several unrelated sounds, including a stapler stapling, a manhole cover settling, a tire going over a curb and a meat locker closing. But this sound comes courtesy of digital audio.

"We lined up the elements in the computer so I could immediately play back five or six sounds synchronized to each other," says digital-sound editor Gene Gearty of the Sound One Corporation in Manhattan. It took him

less than 20 minutes to create the joint effect. Gearty and a handful of other digital enthusiasts are making waves in the staunchly analog world of film-sound production. They are talking bit size and random access in an industry whose stock tools are razor blades and splicing tape.

Digital audio has become the darling of the music industry largely because binary code, unlike an analog signal on vinyl or magnetic tape, does not deteriorate. There are reasons the film industry might become enamored as well. Editing analog sound requires physically cutting a tape; playing several different sounds simultaneously requires synchronizing several machines. In contrast, digital recording "samples" a sound many thousand times per second and stores profiles of the sound in binary code. Editing and synchronizing digital sound is as easy as altering or merging text files on a word processor.

With his digital-editing machine, Gearty has virtually instant access to thousands of sounds in a five-gigabyte optical-disk library and virtually unlimited freedom to vary the balance, volume, reverberation and pitch of the different elements. Yet many film people are eyeing the technology with skepticism. Strangely, the same attributes its champions praise—speed, convenience and sound quality—are the ones that make other sound editors nervous. "There's something to be said for the time it takes to think about a mix or an edit," says Mark Berger, a rerecording mixer at the Saul Zaentz Company Film Center in Berkeley. "And interesting sounds are better than perfect sounds."

The film industry's queasiness may simply be institutional inertia. But other recording technologies have infiltrated film without causing a flap. The industry embraced Dolby noise reduction when its film version was introduced in 1974, and studios routinely spend hundreds of thousands of dollars to upgrade their mixing boards. A switch to digital, however, would necessitate changes not only in equipment but also in the nexus of control of a film's sound. "We end up stepping on a lot of toes," says Skip Lievsay, a digital advocate who supervises the sound on most of Martin Scorcese's films.

Sound in film has traditionally been a collaborative effort involving at least a dozen people and technology that has not changed significantly in the last 20 years. Location recorders capture sound effects and dialogue with a 20-pound, $12,000, allegedly indestructible tape machine called a Nagra. After the film is shot, so-called Foley artists record additional effects (footsteps, car doors slamming, coffee pouring) in synchronization with the film action. Actors and actresses dub lines in the dialogue that were not recorded clearly during shooting. Editors may contribute their own ambience effects (city noise, murmuring crowds, gusts of air), and they may "borrow" sounds from effects libraries as well.

Eventually all sound is transferred to tape deposited on the same kind of sprocketed substrate that carries the optical image. With the help of a manually operated synchronizer, a splicing assemblage and a playback machine, an editor essentially cuts and pastes sounds together to match the cuts that have been made in the film. Mixers then balance the tracks in the separate components of film sound—dialogue, music and effects—before bringing them together for the final mix. The sound on release prints is encoded optically in two quarter-inch tracks running adjacent to the optical image.

A movie soundtrack can have as few as two elements in a dialogue scene or as many as 200 in a riot scene. "There are so many elements involved these days that it's very difficult to maintain any kind of control or even perspective," Lievsay says. "The only way we can deal with it is electronic editing." With digital technology, one person can do the work of three or four editors and a mixer, saving time but also consolidating artistic control. "We're determining what we want instead of paying for expensive mixes," Gearty says. The digital editor may rely less on the director as well, because directors typically do not hear the sound for a film until it is being mixed.

In spite of that control, or perhaps because of it, only a small percentage of film sound today is digitally created or manipulated. And digital technology does not always live up to its potential in film, because so much of the technology is targeted to the music industry. "We have to use hybrid stuff that is not really made for our business," Lievsay says. That is certainly true of the industry's most common digital device, the Synclavier. Designed as a musical instrument in the late 1970's by the New England Digital Corporation, the Synclavier is a kind of glorified synthesizer that not only can make sounds but also can modify any existing sound that is fed into it. The

Computers can give film-sound editors greater artistic control—too much control, some think

***MIXING STUDIO** at Sprocket Systems in San Rafael, Calif., allows workers to synchronize sound and picture and to manipulate sound recorded on as many as 40 separate magnetic tracks. Much of the work currently done on the mixing boards could be superseded by sound editors employing digital technology.*

Synclavier and similar devices have been adopted by astute sound editors who realized that bird warbles, dog barks, gun shots and engine roars could be piped into the keyboard as easily as the sound of a string bass could. Yet even the instrument's fans describe it as a frustrating combination of overkill and oversight.

Combined with optical-disk storage, a Synclavier system can cost anywhere from $100,000 to $500,000, a far cry from the $20,000 needed to equip the conventional editing room. On the other hand, digital machines make possible effects that could not be achieved any other way. In the movie *Who Framed Roger Rabbit?,* for example, a Synclavier helped to construct a sequence in which Daffy and Donald Duck play pianos at breakneck speed. The pianos were real, but the speed was not: on the Synclavier, tempo can be altered without changing pitch.

Even the technology's detractors admit that digital audio is a useful tool for some effects. "For the broad strokes, it's very good," says Maurice Schell, a 23-year veteran of film sound who eschews the technology as a matter of principle. Schell concedes it makes sense to use digital sound for backgrounds, ambience and what he calls "presence": the rush of air, the claustrophobic sound of a tunnel, the din of traffic.

In the meantime, however, the analog competition is improving. The latest Dolby format for analog sound, called SR (for "spectral recording"), extends the low- and high-end frequencies and the dynamic range of noise reduction and is very highly regarded in the sound industry. "With SR, analog does actually sound better than digital," says Tom Scott of Sprocket Systems, a division of Lucasfilm Ltd. In fact, a company launched by Lucasfilm terminated its own "sound droid" digital-editing project in 1987. Now Scott says Lucasfilm is working with New England Digital to improve on the Synclavier design. "We're still waiting for what we think is the maturity of digital storage, retrieval and editing."

Scott and others say the industry's perennial infatuation with state-of-the-art technology makes it likely that all film sound will eventually be digital—from initial recording to delivery in the theater. Before that happens, considerable advances must be made on either side of the editing room. The digital audiotape (DAT) machines sometimes used for location recording are sensitive and unreliable. "If there's a loud noise or if an actor shouts unpredictably, the sound distorts," Lievsay says.

In addition, there is no standard procedure for encoding digital information onto film to make release prints. An evaluation from the Eastman Kodak Company suggests that the minimum permissible bit size for optical recording would be 20 microns, an area that would prohibit transfer of the entire signal to the space allotted for the soundtrack on 35-millimeter film. The Optical Radiation Corporation in Azusa, Calif., is investigating methods for putting digital sound on 70-millimeter film.

All this will be for naught unless the sound systems in theaters improve. Ioan Allen, a vice-president at Dolby Laboratories, Inc., who shared an Oscar with Ray Dolby this year for contributions to film sound, calls the "mediocrity of theaters" the most intractable limitation on both digital and analog technologies. Because of the flawed design of most theaters and their substandard equipment, Allen says the SR format works to his satisfaction in only about a third of the 200 venues in this country equipped to run it. Digital film sound is likely to suffer the same fate until theaters take sound quality as seriously as the film industry does. —*Karen Wright*

MERCURY

MERCURY SABLE WAGON. A special rear suspension was designed to give the Sable wagon the outstanding cornering characteristics of a Sable sedan—quite an achievement in a station wagon. Also, its well-rounded shape manages the wind to help keep the Sable wagon pressed to the road for better control. Even in the corners. Experience the handling of a Mercury Sable wagon at your Lincoln-Mercury dealer today. For more Sable wagon information, call 1-800-822-9292.

SABLE—WINNER OF THE "DRIVER'S CHOICE AWARD" AS BEST STATION WAGON FOR 1989 FROM PBS TV's *MOTORWEEK*.

For larger or smaller wagon needs, see Mercury's Grand Marquis Colony Park or Tracer wagons.

Ford MERCURY LINCOLN

LINCOLN-MERCURY DIVISION Ford Buckle up—together we can save lives. **Quality is Job 1.**

SCIENTIFIC AMERICAN June 1989 Volume 260 Number 6

Toward a New Industrial America

A "bottom-up" study of U.S. industrial performance—from the factory floor to the corporate boardroom—by a distinguished group of experts reveals worrisome weaknesses but also encouraging signs of vitality

by Suzanne Berger, Michael L. Dertouzos, Richard K. Lester, Robert M. Solow and Lester C. Thurow

The U.S. economy is a perplexing mix of strengths and weaknesses. It is now in the seventh year of the longest peacetime expansion in this century. Since the early 1980's large numbers of new jobs have been created, and both unemployment and inflation have remained low. American exports have recently surged (helped by a decline in the exchange value of the dollar), and in late 1988 American factories were operating at close to full capacity.

On the negative side, the trade deficit remains formidable (although it is beginning to shrink). In 1988 the U.S. bought about $120 billion more goods and services from other countries than it could sell overseas. The U.S. automobile and steel industries, which once dominated world commerce, have lost market share both at home and abroad, and newer industries are also struggling. The American presence in the consumer-electronics market, for example, has all but disappeared.

There are other disturbing signs that American industry as a whole is not producing as well as it ought to produce or as well as the industries of other nations have learned to produce. Growth in productivity, a crucial indicator of industrial performance, has averaged only slightly more than 1 percent per year since the early 1970's. Productivity has grown more rapidly in several Western European and Asian nations, and U.S. firms are increasingly perceived to be doing poorly in comparison with their foreign competitors in such key aspects as the cost and quality of their products as well as the speed with which new products are brought to market. In many new fields with broad commercial applications, such as advanced materials and semiconductors, America's best technology may already have been surpassed.

In spite of such disquieting developments, some observers maintain that there is nothing fundamentally wrong with American industry itself. The trade deficit, in this view, is the result not of intrinsic deficiencies in industrial performance but rather of such macroeconomic factors as natural differences in rates of economic growth among countries, fluctuations in currency-exchange rates and the enormous U.S. budget deficit. Then, too, the rise and fall of industries is said to be a normal part of economic evolution; at any given time a certain number of industries are sure to be in decline while others are growing.

Yet if the unfavorable trends in industrial performance are real (and we believe they are), then the U.S. has reason to worry. Americans must produce well if Americans are to live well. The sluggish growth in U.S. productivity is barely sufficient to sustain an improvement in the nation's standard of living. (Real wage rates have in fact hardly increased since the early 1970's.) That, in itself, would be of concern regardless of what is happening in the rest of the world. As it is, the more dynamic productivity performance of other countries is also resulting in a relative decline in the U.S. standard of living. Moreover, because political and military power depend ultimately on economic vitality, weaknesses in the U.S. production system will inevitably raise doubts about the nation's ability to retain its influence and standing in the world at large.

Late in 1986 the Massachusetts Institute of Technology established the Commission on Industrial Productivity (with funding from the Sloan and Hewlett foundations) to determine whether there actually are pervasive weaknesses in U.S. industrial practices and, if so, to identify their causes and formulate a set of recommendations to counter them. Unlike many observers of contemporary U.S. industry, the commission did not view the problem entirely in macroeconomic terms. We believed that we could best contribute to the understanding of the problem by focusing on the

SUZANNE BERGER, MICHAEL L. DERTOUZOS, RICHARD K. LESTER, ROBERT M. SOLOW and LESTER C. THUROW were among the 16 faculty members of the Massachusetts Institute of Technology who formed the M.I.T. Commission on Industrial Productivity. Berger is professor in the political science department, which she also heads. Dertouzos, who was chairman of the commission, is professor of electrical engineering and computer science; he is also director of the M.I.T. Laboratory for Computer Science. Lester is professor of nuclear engineering and served as the commission's executive director. Solow, who was the commission's vice-chairman, is institute professor in the department of economics. He is the recipient of the 1987 Nobel prize in economics. Thurow is professor of economics and management and dean of the M.I.T. Sloan School of Management.

nation's production system: the organizations, the plants, the equipment and the people—from factory workers to senior executives—that combine to conceive, design, develop, produce, market and deliver products.

In keeping with this "bottom-up" approach, the commission began its task by dividing into eight teams, each of which would examine in detail one of eight manufacturing industries: automobiles; chemicals; commercial aircraft; consumer electronics; machine tools; semiconductors, computers and copiers; steel; and textiles. These industries combined account for 28 percent of U.S. manufacturing output and about half of the total volume of manufactured goods traded by the U.S. (exports and imports). American firms in each industry were evaluated for what we have come to call productive performance: their efficiency, product quality, innovativeness and adaptability, as well as the speed with which they put new products on the market. Such factors are not explicitly captured in conventional measures of industrial productivity. Altogether, the commission's teams visited more than 200 companies and 150 plant sites and conducted nearly 550 interviews in the U.S., Europe and Japan.

In choosing to focus on the production system itself, we did not underestimate the importance of the macroeconomic factors that regulate the economy in the large; on the contrary, we could not avoid observing their manifestations time and again as the teams proceeded with their work. It is clear that the nation's productive performance problems will not be solved without some improvement in the economic environment. The reason is that investment—meant here broadly to include not only new plants, equipment and public works but also education, training and research and development—is crucial for productivity, and the economic environment largely determines the level of a nation's investment. Indeed, we believe that the highest priority of U.S. economic policy must be to reduce the huge federal budget deficit, which saps the savings from which investment funds are drawn.

Nevertheless, after two years of study, it seems clear to us that current economic conditions do not fully explain the deficiencies in U.S. industrial performance, nor will macroeconomic policy changes suffice to cure them. The relation of poor product quality to U.S. interest rates and tax policies, for example, seems at best tenuous. The economic environment also does not directly affect the speed with which firms identify and respond to changes in the market and to new technological possibilities. Finally, macroeconomics cannot adequately explain why some U.S. businesses thrive in the very same sectors where others are failing, nor why Japanese manufacturing plants in the U.S. have often achieved better results than comparable American plants.

By looking at what actually takes place in industry—from the shop floor to the boardroom—the commission was able to observe recurring patterns of behavior and to draw certain conclusions about the most important micro-level factors that have adversely affected U.S. industrial performance. To do so the commission worked much like a jury: we assessed the large mass of detailed, diverse and sometimes contradictory evidence that the study teams had collected, ultimately reaching a verdict.

The verdict is that U.S. industry indeed shows systematic weaknesses that are hampering the ability of many firms to adapt to a changing international business environment. In particular, the commission observed six such weaknesses: outdated strategies; neglect of human resources; failures of cooperation; technological weaknesses in development and production; government and industry working at cross-purposes; and short time horizons.

The industry studies revealed two types of outdated strategies that are impeding industrial progress today: an overemphasis on mass production of standard commodity goods and an economic and technological parochialism. Both are holdovers from the unique economic environment that prevailed after World War II. For decades after the war U.S. industry was able to flourish by mass-producing undifferentiated goods principally for its own markets, which were large, unified and familiar. Because firms in most other countries had to rebuild in economies devastated by the war, they could mount no significant competition and were largely ignored by U.S. industry.

Not only did U.S. producers sell their wares primarily to the domestic market, they also drew their technical expertise almost exclusively from U.S. factories and laboratories. Such technological parochialism blinded Americans to the growing strength of scientific and technological innovation abroad and hence to the possibility of adapting foreign discoveries. In the 1950's and 1960's, for example, American steel producers lagged behind Japanese and European steelmakers in adopting such new process technologies as the basic oxygen furnace; later they were again slow to adopt continuous casters and such quality-enhancing technologies as vacuum degassing and oxygen injection. The critical error in many of these cases was the failure to recognize the worth of someone else's innovation.

The American industry of the 1950's and 1960's pursued flexibility by hiring and firing workers who had limited skills rather than by relying on multiskilled workers. Worker responsibility and input progressively narrowed, and management tended to treat workers as a cost to be controlled, not as an asset to be developed.

Training practices in the U.S. have been consistent with that strategy. Workers often receive limited training while on the job; typically it amounts to watching a colleague at work. Even in firms offering organized training programs, in-plant training is usually short and highly focused on transmitting specific narrow skills for immediate application. In other countries we observed a greater inclination to regard firms as learning institutions, where—through education and training—employees can develop breadth and flexibility in their skills and also acquire a willingness to learn new skills over the long term. In a system based on mass production of standard goods, where cost matters more than quality, the neglect of human resources by companies may have been compatible with good economic performance; today it appears as a major part of the U.S.'s productivity problem.

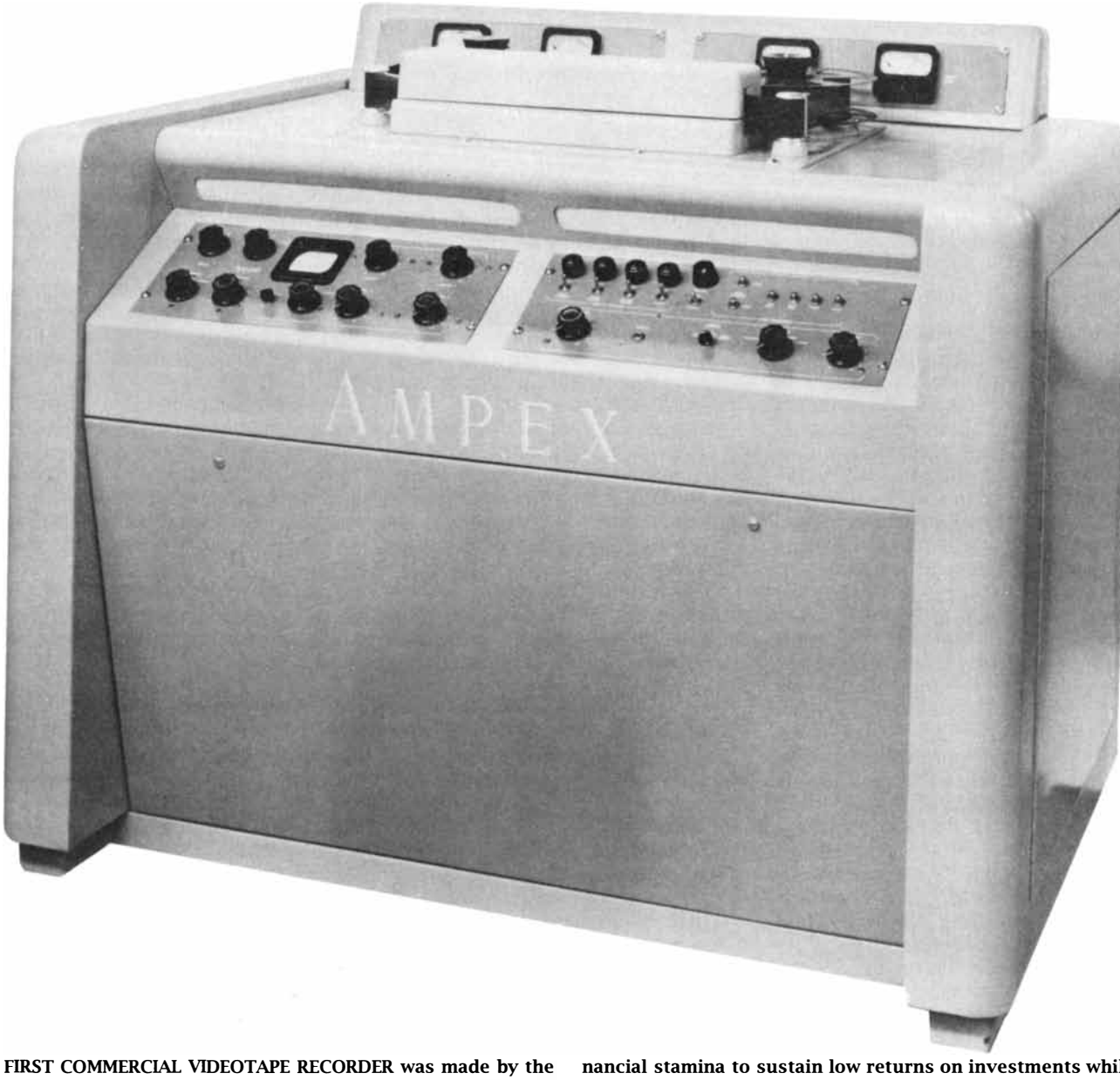
The neglect of human resources in the U.S. actually begins long before young Americans enter the work force. It is in primary and secondary school that they learn the fundamental skills they will apply throughout life: reading, writing and problem solving. Yet cross-national research on educational achievement shows American children falling behind children in other societies in mathematics, science and language attainment at an early age and falling farther behind as they progress through the school years. The school system—from kindergarten through high school—is leaving large numbers of its graduates without basic skills. Unless the nation begins to remedy these inadequacies in education, real progress in improving

the U.S.'s productive performance will remain elusive.

The third recurring weakness of the U.S. production system that emerged from our industry studies is a widespread failure of cooperation within and among companies. In many U.S. firms communication and coordination among departments is often inhibited by steep hierarchical ladders and organizational walls. In addition, labor and management continue to expend valuable resources and energies battling over union organizing.

Suppliers and even customers have also been kept at arm's length by the management of many U.S. companies, in spite of the fact that such vertical linkages can be conduits not only for raw materials and finished products but also for technological innovations and other developments that enhance productivity. These companies are reluctant to share designs, technologies and strategies with either their customers or their suppliers for fear that proprietary information will leak to competitors. Yet by keeping that kind of information to itself, a firm misses the chance to work with its suppliers and customers to improve the products it sells and buys. A similar lack of horizontal linkages—cooperative relations between firms in the same industry segment—has led to a dearth of joint projects in such areas as the setting of common standards and industrial research and development, even when they might have been permitted under the law.

Notwithstanding its spotty performance in the global market in recent years, the U.S. remains the world leader in basic research.



FIRST COMMERCIAL VIDEOTAPE RECORDER was made by the Ampex Corporation in Redwood City, Calif., but no U.S. firms were willing or able to devote the resources to bring unit costs down for sale to retail customers. Ampex concentrated on high-price, high-performance systems; other U.S. firms abandoned the field altogether. Japanese companies had the financial stamina to sustain low returns on investments while perfecting designs and manufacturing processes. The result is that the Japanese now dominate the consumer video-recording market. Moreover, by capitalizing on the profits, technology and economies of mass production built up in that market, they have begun to encroach on the upscale market as well.

THE M.I.T. COMMISSION ON INDUSTRIAL PRODUCTIVITY

Michael L. Dertouzos (Chairman)
Department of Electrical Engineering and Computer Science
Director, Laboratory for Computer Science
Robert M. Solow (Vice-chairman)
Institute Professor
Department of Economics
Richard K. Lester (Executive Director)
Department of Nuclear Engineering
Suzanne Berger
Head, Department of Political Science
H. Kent Bowen
Department of Materials Science and Engineering
Codirector, Leaders for Manufacturing Program
Don P. Clausing
Department of Electrical Engineering and Computer Science
Eugene E. Covert
Head, Department of Aeronautics and Astronautics
John M. Deutch
Provost
Merton C. Flemings
Head, Department of Materials Science and Engineering
Howard W. Johnson
Special Faculty Professor of Management, Emeritus
President, Emeritus
Honorary Chairman, M.I.T. Corporation
Thomas A. Kochan
Sloan School of Management
Daniel Roos
Department of Civil Engineering
Director, Center for Technology, Policy and Industrial Development
David H. Staelin
Department of Electrical Engineering and Computer Science
Lester C. Thurow
Dean, Sloan School of Management
James Wei
Head, Department of Chemical Engineering
Gerald L. Wilson
Dean, School of Engineering

MEMBERS of the Commission on Industrial Productivity were drawn from the faculty of the Massachusetts Institute of Technology. The interdisciplinary group included economists, technologists and experts on organization, management and politics.

Ironically that outstanding success may have diverted attention from "downstream" technological skills in product and process development and production that become progressively more important as new concepts proceed down the path from the laboratory to the marketplace. Simply put, many U.S. firms have been outperformed in the design and manufacture of reliable, high-quality products.

A survey conducted by the International Motor Vehicle Program (IMVP) at M.I.T. found that, despite recent gains, the number of defects reported in the first three months of use was still almost twice as high for cars produced in American plants in 1986 and 1987 as for those from Japanese plants. The commission's automobile study team also learned that American car builders have recently been taking about five years to carry a new design from the conceptual stage to commercial introduction. In contrast, Japanese manufacturers complete the cycle in three and a half years.

Some of the responsibility for the persistent failure to convert technologies quickly into viable, high-quality products lies in the American system of engineering education, which has deemphasized product realization and process engineering since World War II. The professional norms of the American engineering community also assign rather low priority to such essential downstream engineering functions as the testing of product designs, manufacturing and product and process improvements.

Other aspects of the problem can be found in certain practices followed by U.S. industry. For one, many American companies simply do not devote enough attention to the manufacturing process. In a recent comparative study of industrial research and development in Japan and the U.S., Edwin Mansfield of the University of Pennsylvania found that U.S. companies are still devoting only a third of their R&D expenditures to the improvement of process technology; the other two thirds is allocated to the development of new and improved products. In Japan those proportions in R&D expenditures are reversed.

Many U.S. companies also fail to coordinate product design and the manufacturing process. It has been standard practice for design engineers to end their involvement with a new product once they have conceived its design. They hand over the design to manufacturing engineers, who are then supposed to come up with a process for the product's manufacture. This compartmentalization of tasks has led to serious problems. Product-design groups often neglect manufacturing considerations, making it harder to come up with a manufacturing process.

The Proprinter project of the International Business Machines Corporation is an impressive example of what can be achieved when product designers are brought together with manufacturing engineers and research scientists. Charged with designing a new computer printer that has fewer component parts and no springs or screws (which increase assembly time and decrease reliability), a multidisciplinary IBM design team came up with a product having 60 percent fewer parts than its predecessor. (Ironically because an individual assembly worker could put the printer together in three and a half minutes, the highly automated and expensive assembly plant that had been built to make it was largely rendered superfluous.)

Multifunctional design teams and an orientation toward simplicity and quality at the design stage have been a long-standing fixture of Japanese industry and have contributed to its comparative advantages in quality and productivity. The IMVP survey showed that Japanese-designed automobiles retain their quality advantage even when they are assembled in American factories, which implies that the Japanese automotive engineers had incorporated quality-enhancing features into the design itself.

American companies also have often lagged behind their overseas competitors in exploiting the potential for continual improvement in the quality and reliability of products and processes. The cumulative effect of successive incremental improvements in and modifications of established products and processes can be very large; it may even outpace efforts to achieve technological breakthroughs.

The federal government deserves part of the blame as well for the technological weaknesses in development and production. Whereas the governments of most other industrial nations have purposefully promoted research and technology for economic development, U.S. policy for science and technology has traditionally focused on basic research. The commercial development and application of new technologies have for the most part been considered to be the responsibility of the private sector.

To be sure, the Department of Defense, the National Aeronautics and Space Administration and other federal agencies have invested heavily in technology development. Indeed, about 46 percent of all U.S. research and development is sponsored by the Government. Those expenditures are

usually in the areas of defense and space activities or in other specific Governmental missions, however. In such cases commercial applications of the resulting technology are considered secondarily, if at all. Furthermore, there are indications that defense R&D, which accounts for almost two thirds of all federal R&D spending, is becoming less relevant to the needs of the civilian market.

More generally the lack of a common agenda between government and industry has produced negative effects across broad stretches of the U.S. economy. Some observers, for instance, have blamed the collapse of the consumer-electronics industry in part on the federal government's failure to enact or implement tariffs and import quotas as well as to amend or enforce antidumping and antitrust laws. Yet while some see the problem as too little government support for key industries, others see it as too much government support for inefficient producers.

The evidence gathered from the commission's industry studies was similarly mixed regarding the charge that too much government intervention, particularly in regulating the environment and occupational safety, has put U.S. companies at a disadvantage in relation to foreign competitors. Where problems have arisen, the fault tended to lie in the nature of the regulatory process rather than in the strictness of the regulations themselves. Indeed, many European countries as well as Japan now have environmental and occupational-safety laws in many areas that are at least as strict as those in the U.S.

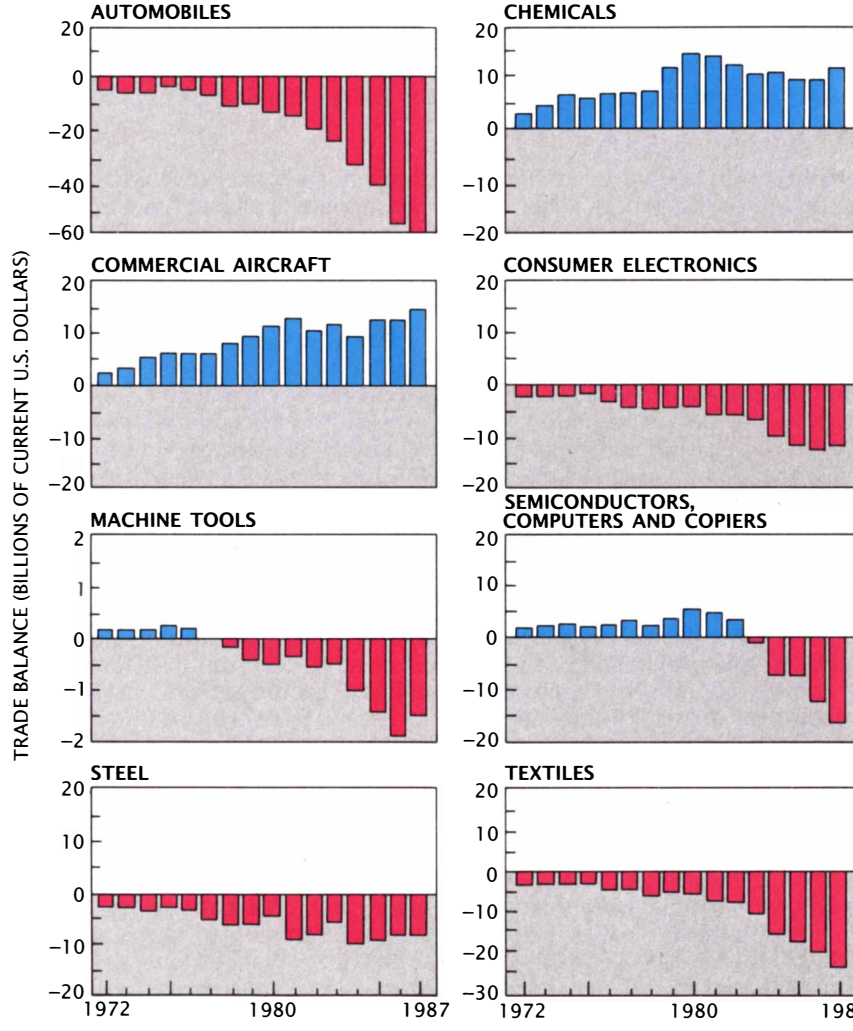
The issue, then, is not simply whether there is too much government or too little. What is clear to the commission, however, is that a lower level of cooperation between government and business exists in the U.S. than it does in the countries of American firms' major foreign competitors and that the frequency with which government and industry find themselves at cross-purposes is a serious obstacle to strategic and organizational change in individual U.S. firms.

American industry has also been handicapped by shrinking time horizons and a growing preoccupation with short-term profits. There have been many recent instances in which U.S. firms have lost market share to overseas competitors despite an early lead in technology or sales, or both. Often these firms effectively cede a potential market by not "sticking to their knitting"; instead, they diversify into activities that are more profitable in the short run.

The development of the videocassette recorder provides an exemplary case. Video-recording technology was first developed in the U.S., but the early machines were complex and expensive and suitable only for industrial and professional applications; many years of further development were needed to create low-cost, highly reliable products for the mass-consumer market. No American manufacturer was willing or able to spend the time and money, but several Japanese manufacturers were. The Japanese are now virtually unchallenged as makers of the most important single product in the consumer-electronics market.

Why are U.S. firms less willing than their rivals to live through a period of heavy investment and meager returns in order to build expertise and secure a foothold in a new market? Is it that American managers are incapable of looking as far ahead as their foreign counterparts? Or are they forced by external circumstances to focus on the short term, even though they realize that it is not in their firm's best interest to do so? Or might it be that a short-term focus is actually in the best interest of the firm but not of the U.S. economy as a whole?

Some observers argue that the higher cost of capital in the U.S. compared



BALANCE OF TRADE (the total value of exports less the total value of imports) for the eight key manufacturing industries studied by the commission reflects the industries' general condition in the U.S. A positive balance means that more of an industry's products are sold overseas than are exported to the U.S. by foreign rivals; a negative balance implies the converse. An industry's trade balance is affected by the performance of its firms with respect to product cost and quality, service and the speed of response to new technological and market opportunities. Macroeconomic conditions—particularly currency-exchange rates—can also affect the balance.

NATURE OF U.S. SECURITIES MARKET has contributed to the short time horizons of American businesses. Managers of mutual and pension funds, which own a large and growing share of the capital of U.S. firms, tend to turn over their fund's stockholdings rapidly in an effort to maximize the current value of the investment portfolio. Such a strategy undervalues long-term development and investment policies of U.S. firms.

with its cost in Japan is the overriding reason for the different time horizons of firms in the two countries. Certainly the cost of capital is important, but we think that other factors are also important.

The nature of the institutions that influence the supply of capital may affect investment decisions at least as much as the cost of capital. A large and growing share of the capital of U.S. firms is owned by mutual funds and pension funds, which hold assets in the form of a market basket of securities. The actual equity holders, the clients of the funds, are far removed from managerial decision making. The fund managers also have no long-term loyalty to the corporations in which they invest and have no representation on their boards. (Indeed, legislation prohibits their participation in corporate planning.)

Although some fund managers do invest for the long term, most turn over their stockholdings rapidly in an effort to maximize the current value of their investment portfolio, since this is the main criterion against which their own performance is judged. Firms respond to this financial environment by maximizing their short-term profit in the belief that investment policies oriented toward the long term will be undervalued by the market and thus leave them vulnerable to a take-over.

At the same time senior executives are also motivated to maintain steady growth in earnings by their own profit-related bonus plans and stock options. A chief executive whose compensation is a strong function of his or her company's financial performance in the current year will naturally stress short-term results.

Explanations that cite the cost of capital and the sources of financing all tend to depict corporate managers as victims of circumstance, forced by external conditions into a short-term mind-set. Yet Robert H. Hayes and the late William J. Abernathy of the Harvard Business School have argued that executive ranks have come to be dominated by individuals who know too little about their firm's products, markets and production processes and who rely instead on quantifiable short-term financial criteria. These modern executives are more likely to engage in restructuring to bolster profits than to take risks on technological innovation.

As part of its work the commission sought to find not only patterns of weakness in U.S. industry but also patterns of change that are common to successful U.S. firms—firms that are doing well in the international arena. Indeed, we probably learned as much from what such "best practice" firms are doing right as from what many other U.S. firms are doing wrong.

In particular, we found that successful firms emphasize *simultaneous* improvements in quality, cost and speed of commercialization. Whereas other firms often trade off one dimension of performance against another, only the best companies have made significant improvement in all three. To gauge progress, one common practice among the successful firms is to emphasize competitive benchmarking: comparing the performance of their products and work processes with those of the world leaders. At the Xerox Corporation, for example, quality improved by an order of magnitude over the past decade after the company instituted detailed comparison tests of Xerox copiers and competing Japanese models.

In addition, the best-practice firms we observed are developing closer ties to their customers. These ties enable companies to pick up more detailed signals from the market and thus to respond to different segments of demand. They also increase the likelihood of rapid response to shifts in the market. Even high-volume manufacturers have combined a continuing emphasis on economies of scale with a new flexibility, reflected in shorter production runs, faster product introductions and greater sensitivity to the diverse needs of customers.

Closer and more tightly coordinated relations with suppliers were also observed among the best-practice firms. In some cases, better coordination with suppliers has been achieved through the coercive power of market domination, in others by new forms of cooperation and negotiation. No matter how it comes about, coordination with external firms is crucial in cutting inventories (and thereby costs), in speeding up the flow of products and in reducing defects.

For example, Greenwood Mills, Inc. (a textile company specializing in the production of denim), brought down its inventory radically over two years, even as sales doubled. To achieve those results the company tightened up its own operations and at the same time negotiated new arrangements with suppliers, who now deliver on a just-in-time basis. In exchange, Greenwood Mills halved the number of its suppliers, leaving itself more vulnerable to price hikes but gaining the advantages of closer collaboration.

Most thriving firms in the U.S. have also realized that business strategies based on throwing new hardware at performance problems are unlikely to work. They have instead learned to integrate technology in their manufacturing and marketing strategies and to link them to organizational changes that promote teamwork, training and continuous learning. In the general-

ly depressed domestic apparel industry, firms such as the Model Garment Company and Levi Strauss & Company are succeeding by investing heavily in information technologies that allow them to fill orders very rapidly and reduce inventory levels.

In virtually all successful firms, the trend is toward greater functional integration and lesser organizational stratification, both of which promote quicker product development and increased responsiveness to changing markets. The Ford Motor Company was the first U.S. automobile company to experiment with cross-functional teams to speed the development and introduction of a new model. The product-development team for the Taurus model included representatives from planning, design, engineering, manufacturing and marketing. The specialists worked simultaneously rather than serially.

Flattening steep organizational hierarchies goes hand in hand with dismantling functional barriers. A flatter hierarchy generally enhances organizational flexibility. It also promotes closer relations with customers: a customer with a problem can speak directly with the group that has responsibility for the product instead of having to go through a sales department. In leaner, less hierarchical organizations the number of job categories at each level is reduced, and the responsibilities associated with particular jobs are broadened.

At the Chaparral Steel Company, for instance, there are almost 1,000 employees, and yet there are only four job levels. Production workers are responsible for identifying new technologies, training, meeting with customers and maintaining equipment. Foremen and crews install new equipment. Security guards are trained as emergency medical technicians, and they update computer records while on their shift.

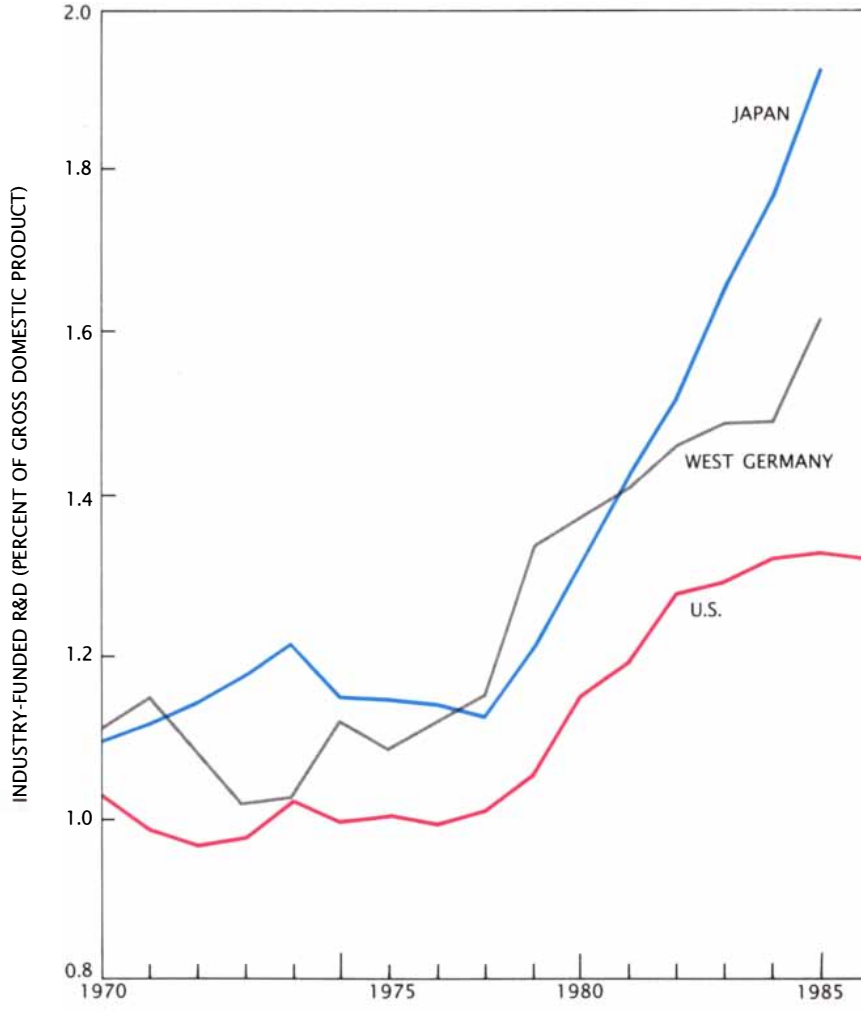
An essential ingredient for greater worker responsibility and commitment is continual training. Large companies such as IBM have the resources to train their own workers. Having lower labor turnover, they also have more incentive to invest in training, because they are more likely to capture the benefits of that investment. Smaller companies do tend to draw more heavily on outside institutions for training, but there is often a major internal component as well.

The Kingsbury Machine Tool Corporation once built dedicated equipment for vehicles; it has since successfully converted to building computer-controlled machines and production lines for flexible manufacturing. Under the old regime the primary demand on the work force was for mechanical skills, but the new product line requires workers with some knowledge of computers. To retrain the employees, the company provided everyone—from janitors to vice presidents—with computers to use at work or at home and offered classes to employees and their families.

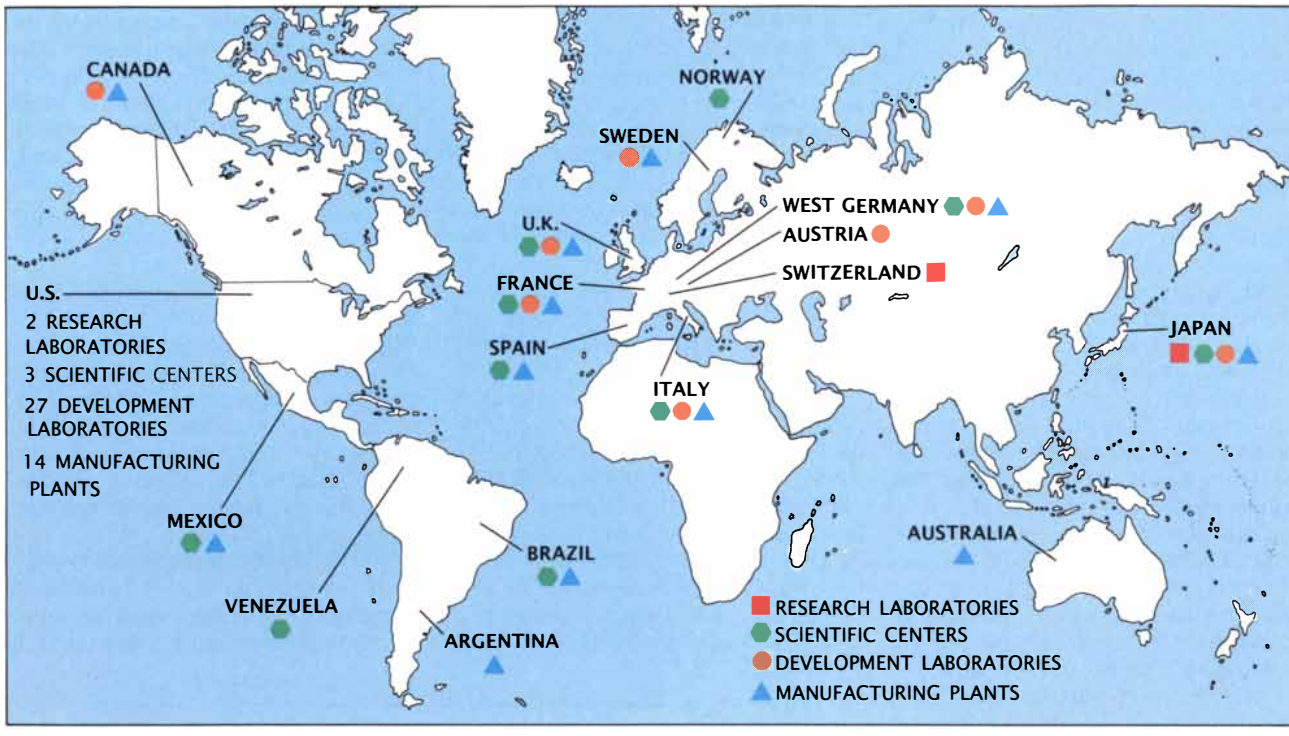
Although an increasing number of American companies are recognizing what it takes to be the best in the world, many U.S. firms have not yet realized that they will have to make far-reaching changes in the way they do business. They will need to adopt new ways of thinking about human resources, new ways of organizing their systems of production and new approaches to the management of technology. What distinguishes the best-practice firms from the others in their industries is that they see these innovations not as independent solutions but as a coherent package. Each change for the better reinforces the others, and the entire organization is affected by them.

Of course, today's best practices will surely not remain the best forever. The nature of industrial competition is changing rapidly, and new challenges will undoubtedly emerge. The commission identified three major and pervasive long-term trends that will have broad implications for the future productive performance of U.S. firms.

First, economic activity will continue to become more international. A company's ownership, location, work force, purchases and sales are all



INDUSTRY-FUNDED RESEARCH AND DEVELOPMENT has grown more slowly in the U.S. than in Japan and West Germany. Total spending on R&D as a percentage of gross domestic product, however, is about the same in the three countries, because the difference in industry-funded R&D is made up in the U.S. by federal funding of R&D.



INCREASING INTERNATIONALIZATION of the U.S. economy has led the International Business Machines Corporation to establish numerous research laboratories, scientific centers, development laboratories and manufacturing plants in foreign countries. Such a geographic distribution helps the corporation remain abreast of technological advances throughout the world. IBM also maintains sales offices in more than 100 countries in order to keep in closer contact with its customers.

spreading beyond the boundaries of the nation in which it originated. A growing number of countries will acquire the capacity to produce and to export sophisticated goods and services. Many of these emerging economies have labor costs even lower than those of Taiwan and South Korea and far lower than those of the U.S., Japan and Europe.

Second, partly because of internationalization and partly because of rising incomes around the world, markets for consumer goods and intermediate goods are becoming more sophisticated. Markets are also becoming more segmented and specialized; not everyone is prepared to accept the same product designs and specifications.

Third, we expect the rapid pace of technological change to continue. Particularly rapid progress seems likely in information technology, materials science and engineering, and biotechnology. Information technology has already permeated nearly every facet of the production of goods and delivery of services, and we expect it to affect the business environment in a number of ways in the future.

The obvious implication of these three trends is that U.S. firms will not be able to compete on the basis of cost alone. The future of U.S. industry lies in specialized, high-quality products; standard commodities will be made in the U.S. only if their production is extraordinarily capital-intensive and technologically advanced. At the same time, competition among U.S., Japanese and European firms in markets for high-value-added products will become increasingly fierce.

Indeed, the convergence of future consumer preferences, market forces and technological opportunities may lead in some industries to the introduction of "totally flexible" production systems. In such systems custom-tailoring of products to the needs and tastes of individual customers will be combined with the power, precision and economy of modern production technology.

In a market economy such as that of the U.S., individual firms have the primary responsibility to correct past problems and find ways to compete successfully in the future. Yet for the U.S. to achieve an economy marked by high productivity growth, all sectors—business, government, labor and educational institutions—will need to work together. Based on its study of current weakness and best practices in American industry, as well as its forecast of long-term trends, the commission believes that five interconnected imperatives should form the core of any such national effort.

First, the U.S. needs to invest more heavily in its future. This means investment not only in tangible factories and machinery but also in research and, above all, in human capital. At the macroeconomic level, as noted earlier, bringing the budget closer into balance should take high priority. In order to encourage firms to develop the necessary outlook for long-term investments, American economic policy should also favor increasing productive investment over private consumption through an approach that combines a more expansionary monetary policy with a fiscal policy that taxes consumption more heavily than savings or investment. Such policies can increase the supply of capital to business. Tax and credit legislation making it harder and more expensive to raise large sums of money for takeovers and buy-outs is additionally needed. Government must also work with industry and academia to ensure not only that investment continues strongly in basic research but also that

it expands in the direction of productive manufacturing technologies.

Public resources should be allocated not only to improve the existing economic infrastructure (roads, airports, harbors and the like) but also to invest in new kinds of infrastructure. For example, we think that the time is right for American business and Government to begin developing a national information infrastructure, which would eventually become a network of communication highways as important for tomorrow's business as the current highway network is for today's flow of goods.

The most important investment in the long run is in the nation's schools. A better basic education will be crucial to the technological competence that will be required to raise the productivity of U.S. industry. Without major improvements in primary and secondary schooling, no amount of macroeconomic fine-tuning or technological innovation will yield a rising standard of living.

The second major imperative, closely related to the first, is to develop a new "economic citizenship" in the workplace. The effective use of modern technology will require people to develop their capabilities for planning, judgment, collaboration and the analysis of complex systems. For that reason learning—particularly through on-the-job training programs—will acquire new importance.

Greater employee involvement and responsibility will be needed to absorb the new production technologies. Companies will no longer be able to treat employees like cogs in a big and impersonal machine. If people are asked to give maximum effort and to accept uncertainty and rapid change, they must be full participants in the enterprise rather than expendable commodities. Just as important as job security is a financial stake in the long-term performance of the firm. We see in this combination of technological and organizational change an unprecedented opportunity to make jobs more satisfying and rewarding for workers at all levels of a firm.

Third, the U.S. needs to make a major commitment to mastering the new fundamentals of manufacturing. Manufacturing, as we use the term here, encompasses a great deal more than what happens on a production line. It includes designing and developing products as well as planning, marketing, selling and servicing them. Global competition, changing markets and modern technologies are transforming virtually every phase of the production system.

Managers who are detached from the details of production will lose the competitive battle to managers who know their business intimately. Manufacturability, reliability and low cost should be built into products at the earliest possible stages of design. Innovation must be applied to process development as intensively and creatively as it is now applied to product development. Corporate management and financial institutions must work together to develop indicators that better reflect how well companies are actually doing in developing, producing and marketing their products than do short-term financial measures such as quarterly earnings. New measures might include indicators of quality, productivity, product-development time and time to market.

Fourth, Americans should strive to combine cooperation and individualism. The nation's culture has traditionally emphasized individualism, often at the expense of cooperation. Yet in the best U.S. companies (as in other societies), group solidarity, a feeling of community and a recognition of interdependence have led to important economic advantages.

To this end, steep organizational hierarchies, with their rigidity and compartmentalization, should be replaced with substantially flatter organizational structures that provide incentives for communication and cooperation among different corporate departments. Companies should put less emphasis on legalistic and often adversarial contractual agreements; they should promote business relations based on mutual trust, common goals and the prospect of continuing transactions over the long run. Management must also accept workers and their representatives as legitimate partners in the innovation process. Both individual and group efforts need recognition and reward.

Americans should think of cooperation among economic entities as a way of overcoming the defects of the market, which often undersupplies collective factors essential to economic success. Cooperative efforts can take the form of research consortiums, joint business ventures, partnerships with Government and standard-setting committees. (To be sure, such arrangements might lead producers to combine forces in order to exploit the consumer. Now and in the future, competition from imports will no doubt provide some protection from domestic monopolies. Still, a little vigilance would help too.)

Fifth, to compete successfully in a world that is becoming more international and more competitive, Americans must expand their outlook beyond their own boundaries. They must gain knowledge of other languages, cultures, market customs, tastes, legal systems and regulations; they will need to develop a new set of international sensitivities.

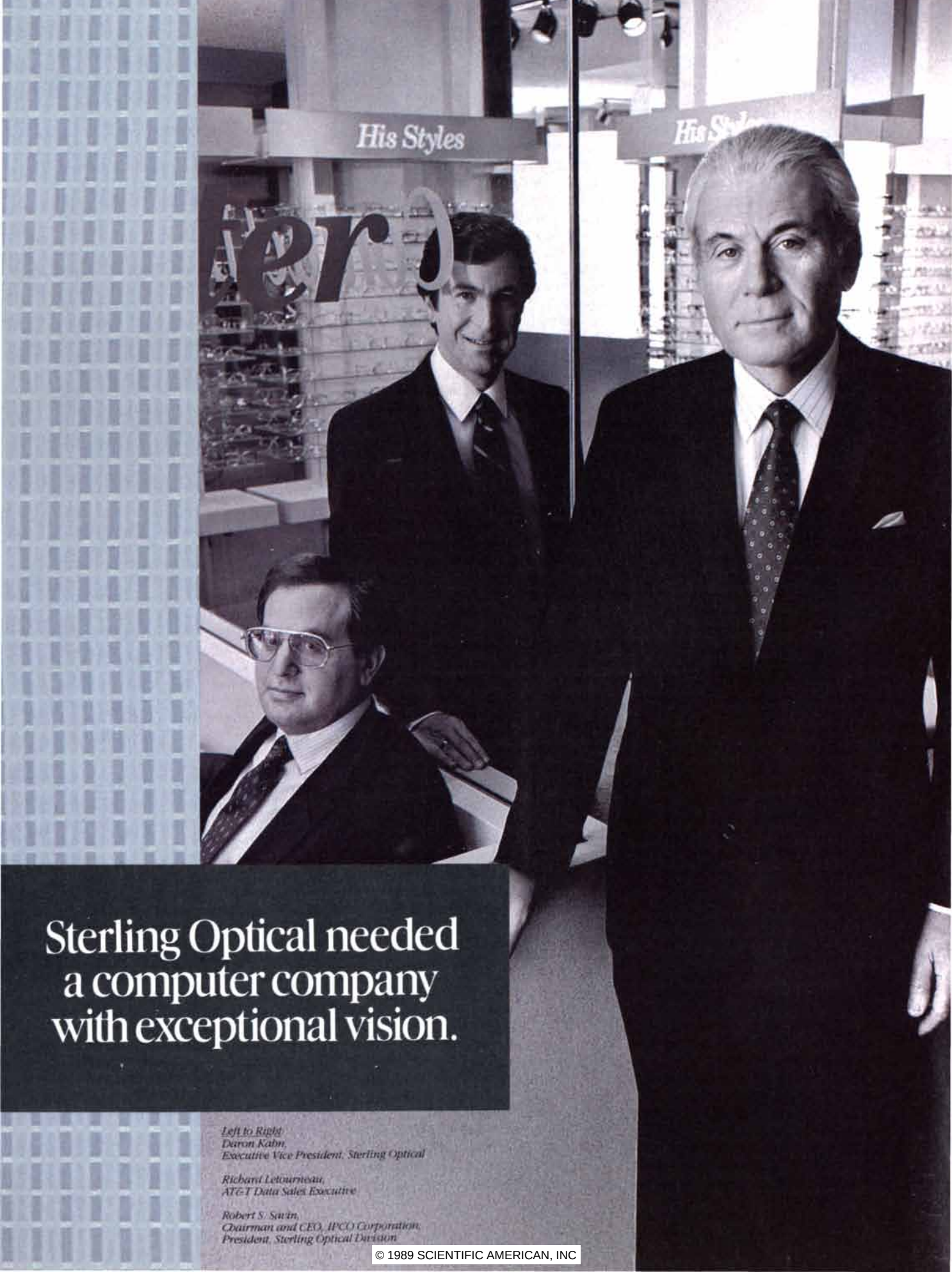
Cost considerations will increasingly dictate whether materials and components are best procured at home or abroad. It follows that not only a company's marketing division but also its purchasing agents and production managers will have to be knowledgeable about global conditions. Shopping internationally should go beyond the buying of raw materials and off-the-shelf products to the adoption of effective practices and technologies—wherever they happen to be found.

Americans need to understand that the world they live in has changed. The effortless economic superiority that the U.S. enjoyed in the aftermath of World War II has gone. Strong economic cultures now exist across both the Atlantic and Pacific oceans. The U.S. has much to learn from the rest of the world. Indeed, the rest of the world will force changes in some of the most cherished American operating procedures and assumptions, if the U.S. is to continue to have a standard of living second to none. What Americans must do is determined decreasingly by what they wish to do and increasingly by the best practices of others.

Implementing these five imperatives will not be easy. In many cases, fundamental changes in attitude will be necessary. Just accepting the need for a sense of common purpose—a shared national goal—may require the biggest attitudinal change of all. The commission believes that if industry, government and the educational system in this country unite in steadfast pursuit of these basic imperatives the next generation of Americans will live in a nation moving into the 21st century with the same dynamism and strength that made it a world leader a generation ago.

FURTHER READING

MADE IN AMERICA: REGAINING THE PRODUCTIVE EDGE. Michael L. Dertouzos, Richard K. Lester, Robert M. Solow and the MIT Commission on Industrial Productivity. The MIT Press, 1989.



Sterling Optical needed a computer company with exceptional vision.

Left to Right
Daron Kahn,
Executive Vice President, Sterling Optical

Richard Letourneau,
AT&T Data Sales Executive

Robert S. Swain,
Chairman and CEO, IPCO Corporation,
President, Sterling Optical Division

Chesterfield, Missouri
February 28, 1989

Sterling Optical, one of the leading optical chains in the country, was looking for a way to get a jump on their competitors in the retail eyeglass and eyecare marketplace. Sterling Optical's Bob Savin, Daron Kahn and AT&T's Rich Letourneau discuss how AT&T developed an open systems-based, retail point-of-sale system to network their 250 stores nationwide.

Sterling: We were getting frustrated by the delays we were facing with the old way of doing things. Our customers were unhappy, we were unhappy, and headquarters wasn't able to keep up with the volume of requests. We knew a highly integrated distributed networked computing solution was the only way to go.

AT&T: And you wanted to preserve the investments you'd already made in applications software—especially those on the System/38.

Sterling: Absolutely. All our RPG-coded customer records and our inventory control system are stored in the database at headquarters. The retail business is demanding—you can't afford to wait to implement brand-new technology. Plus, we don't have a very large computer programming department, and in order to develop, modify, or change corporate programs on the System/38, it's a big effort, and it takes a lot of time.

AT&T: That's why we built our multiuser platform based on the UNIX® System V operating system. We gave you the tools to develop new applications and get them into the mainstream of your business quickly.

Sterling: It made sense. We were able to keep our hardware and software *and* install AT&T 6386 WorkGroup Systems in our stores. Now, on a daily basis, we know what's selling and what's not. We even included an employee time and attendance system that feeds into our existing payroll system.

AT&T: And you've kept your system options open. You can modify any part of the system at any time. Like when you added the automatic pricing software.

Sterling: Right. I think what we like most about the system is its simplicity. Despite the complexity of the information handled, it really gives us easy access to our information. We need that to improve the profit potential of our business, and to maintain our lead in the industry.

AT&T: The system also provides investment protection for what you have today, and a gradual growth path to what you'll need tomorrow.

Sterling: One of the most striking things was AT&T's commitment to service. We came from an environment where it wasn't uncommon to be down for two to three or four days, waiting for equipment to be shipped or repaired.

AT&T: Our message was simple: AT&T wants your business.

Sterling: You were here working as much as we were. And you really listened to us. Of all the vendors we spoke to, you gave us the best proposal, the best equipment, and the best price. In fact, we're so excited about the new system, we wrote it up in our company newsletter.

AT&T: I heard! Can we get a copy?

Sterling: Sure. The only catch is, you have to read it from across the room with one eye closed.

The Sterling Optical Computer Solution:

THE CHALLENGE:

Create a point-of-sale computer system specific to the optical retail business for the 250-store nationwide Sterling Optical chain.

THE SOLUTION:

AT&T created a UNIX System V-based network of AT&T 6386 WorkGroup Systems and AT&T 605 terminals chain-wide. At headquarters, an AT&T 3B2/500 gathers orders and information from each store daily. The system allows a smooth link between the 3B2/500 computer and the existing IBM System/38. A custom INFORMIX database management system forms the heart of the applications software.

THE RESULT:

Sterling Optical customers are getting better service. All store locations will be using one standard point-of-sale system that provides increased productivity, a friendlier working environment, and greatly enhances Sterling Optical's competitiveness in the industry.

Call your AT&T Account Executive, AT&T Authorized Reseller, or 1 800 247-1212, Ext. 530.

Your Computing Systems And Networking Solutions Company

Hidden Earthquakes

Large earthquakes can take place not only on faults that cut the earth's surface but also on "blind" faults under folded terrain

by Ross S. Stein and Robert S. Yeats

An erupting volcano makes its role as a conduit for magma and ash from the interior of the earth unmistakable. The connection between a scarp face and the active fault that sculpted it is less obvious; in fact, the association between earthquakes and such ridges was first made only a century ago. Gently rolling or folded terrain is perhaps the least forbidding landscape, evoking slumber rather than danger. In Matthew Arnold's words:

To a boon southern country he is fled,
And now in happier air,
Wandering with the great Mother's train divine . . .
Within a folding of the Apennine.

But Italy's northern Apennine mountains—one of the many fold belts that girdle the globe—are not still. Rather these mountains are built by earthquakes along faults hidden well under the landscape. Such earthquakes, which unlike their more familiar counterparts do not rupture the earth's surface, we term "surface-folding" earthquakes, or "fold" earthquakes for short. They have only recently become the object of study by investigators.

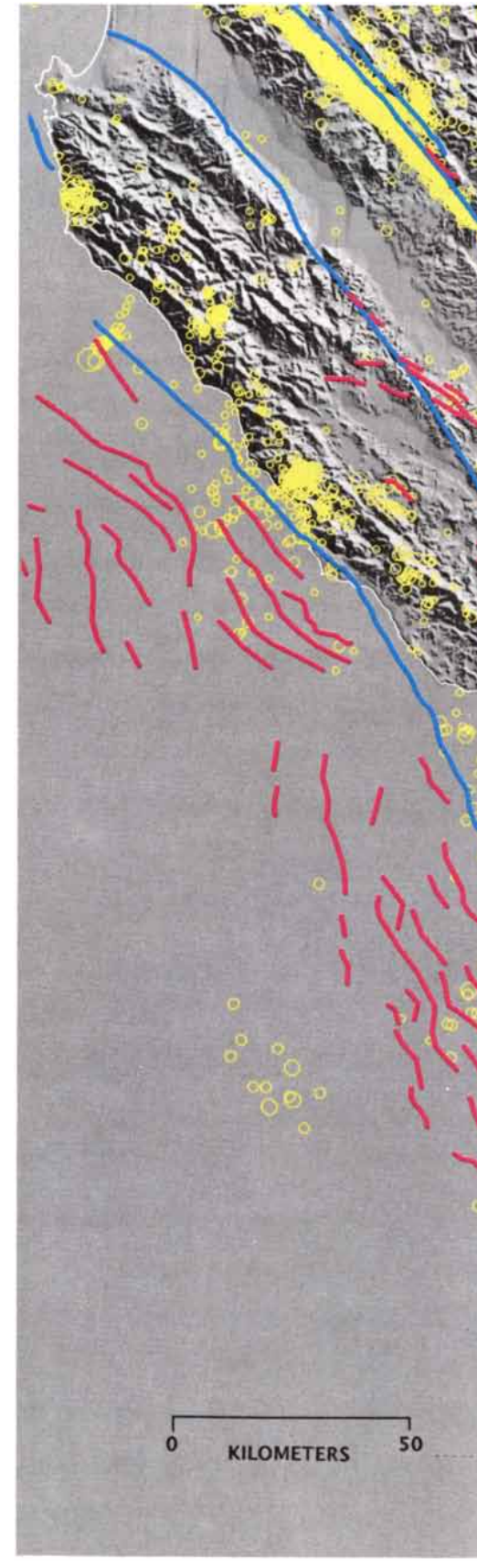
The fundamental premise of earthquake-hazard reduction in the U.S. is that most earthquakes take place along active faults that suddenly fail and slip. Large earthquakes take place along large faults, and large faults, it is widely held, cut the earth's surface. Geologists recognize faults that have displaced young surface deposits as having recently been active and deem them the most likely to rupture again.

This reasoning has yielded profound insights into earthquake behavior, making it possible to situate critical facilities—such as power plants and dams—away from active fault sites, to identify sites with high seismic potential and to make probabilistic forecasts of earthquake size and frequency [see "Predicting the Next Great Earthquake in California," by Robert L. Wesson and Robert E. Wallace; SCIENTIFIC AMERICAN, February, 1985].

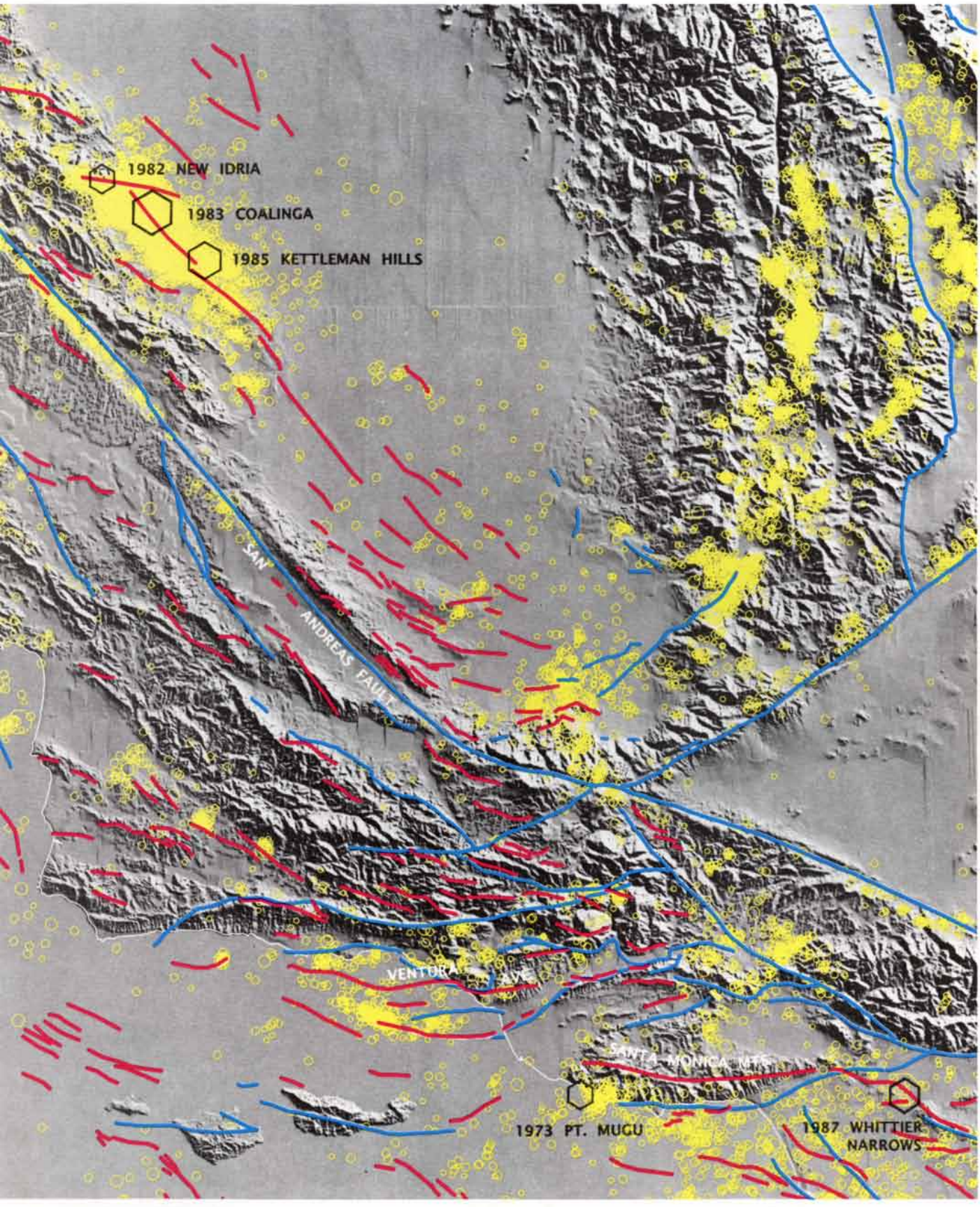
Yet in California most small earthquakes do not occur along faults that cut the earth's surface. Only 50 years ago this fact was ascribed to an inadequate number of seismometers. Now, however, 700 seismometers are in operation in California, and coverage is more complete than in any other area of the world with the possible exception of Japan. Certainly many microearthquakes take place along active faults, such as parts of the San Andreas. But a greater number do not correspond to any known surface fault.

Many of the earthquakes that are not associated with surface faults occur under folds—geologic structures formed when layered sediments are buckled upward in a broad arch called an anticline [*see illustration on page 50*]. The presence of an anticline reflects crustal compression as two moving tectonic plates collide, in much the same way a carpet wrinkles when pushed across the floor. That so many small earthquakes nucleate on faults hidden under folds is remark-

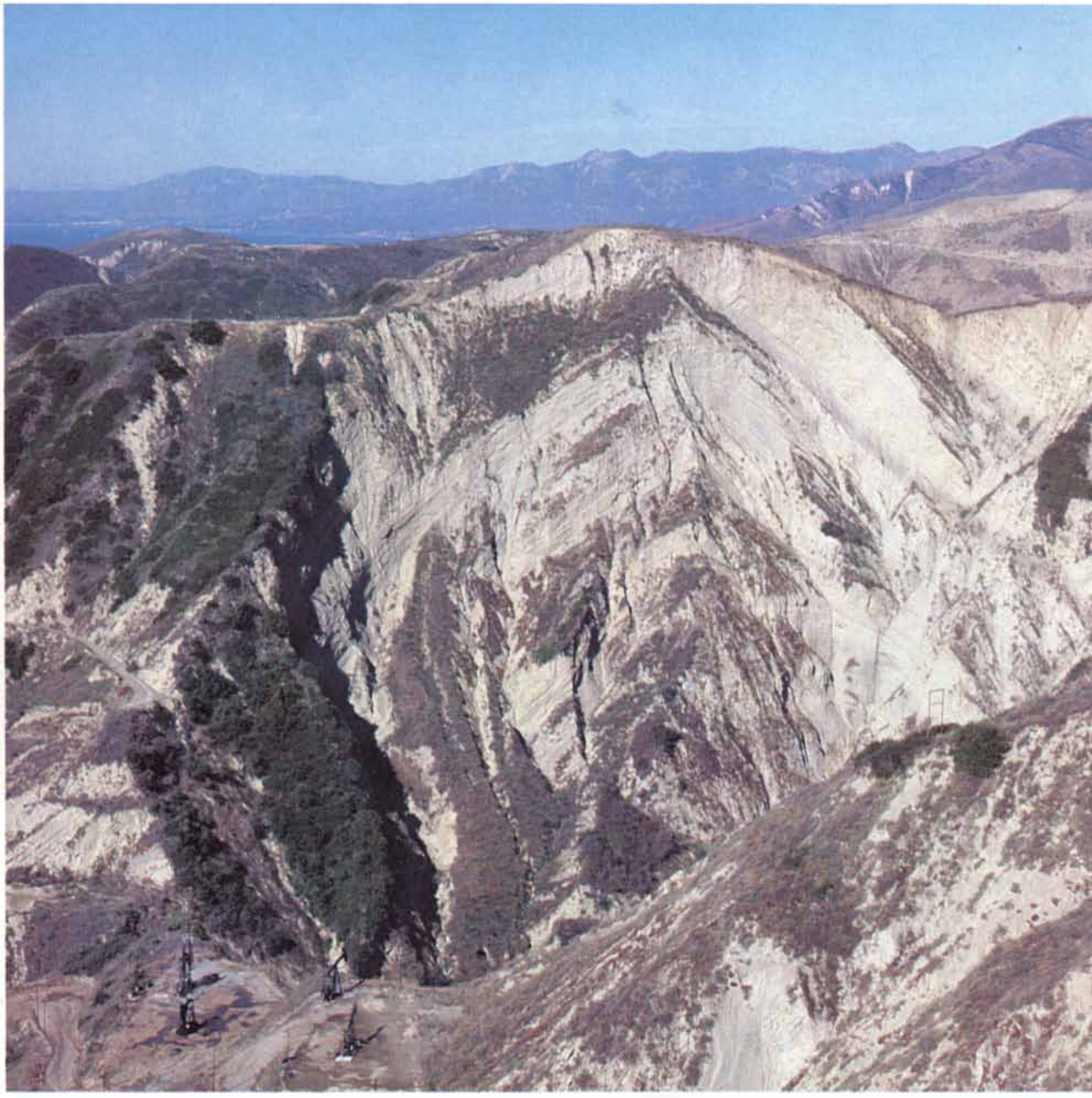
ROSS S. STEIN and ROBERT S. YEATS collaborate on the study of earthquakes because, as Stein puts it, "two heads are better than none." Yeats is professor of geology at Oregon State University and past chair of the department. He received a Ph.D. from the University of Washington in 1958 and then went to work as a petroleum geologist for the Shell Oil Company. In 1967 he joined the faculty at Ohio University and in 1977 went to Oregon. For relaxation he plays the piano. Stein received a B.Sc. at Brown University in 1975 and a Ph.D. at Stanford University in 1980. In 1981 he took his present position as a geophysicist at the U.S. Geological Survey in Menlo Park, Calif. Since 1987 he has also been an editor of the *Journal of Geophysical Research*. For relaxation, Stein says, he eats lasagna baked by Yeats.

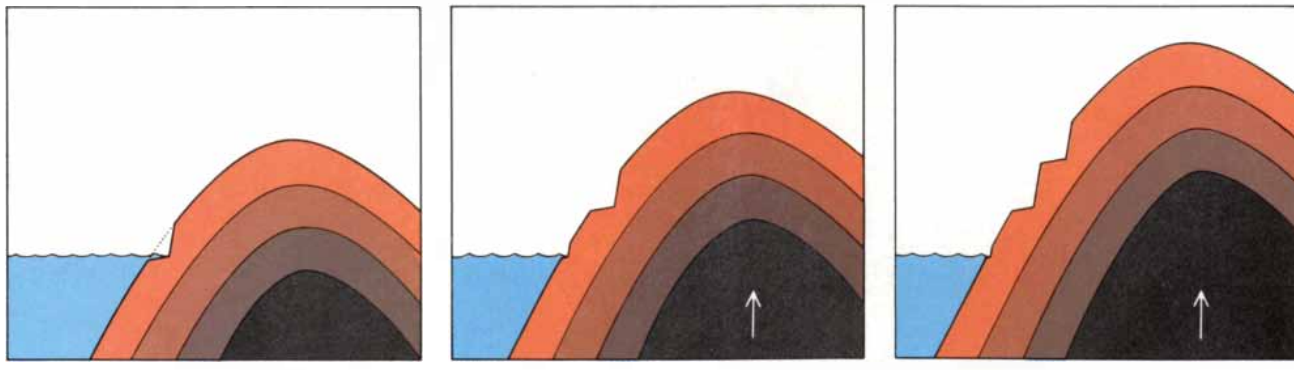


EARTHQUAKES in southern California from 1980 to 1986 are displayed on this topographical map. The yellow circles represent the 23,000 shocks that regis-



tered between 1.5 and 6.5 on the moment-magnitude scale M. Many of the earthquakes are concentrated along surface faults (*blue lines*), in particular the San Andreas. Others, however, appear to be associated with folds (*red lines*). The association is explicable if folds mark the location of active "blind" faults that do not extend to the surface. Recent earthquakes (*hexagons*) that have uplifted folds confirm this hypothesis. No inventory of active folds has yet been made in the U.S.





VENTURA AVENUE ANTICLINE near Ventura, Calif., is one of the world's fastest-growing folds—geologic structures in which layered sediments have buckled upward to form a broad arch. The buckled layers are visible in the photograph (*top*). Nine terraces have been cut into the anticline's flank by storm waves. Carbon dating of seashells shows that the terraces' ages range between 1,800 and 5,600 years. Because sea level has not varied substantially over that time, the simplest explanation is that the terraces have been uplifted by earthquakes recurring about every 600 years (*bottom*).

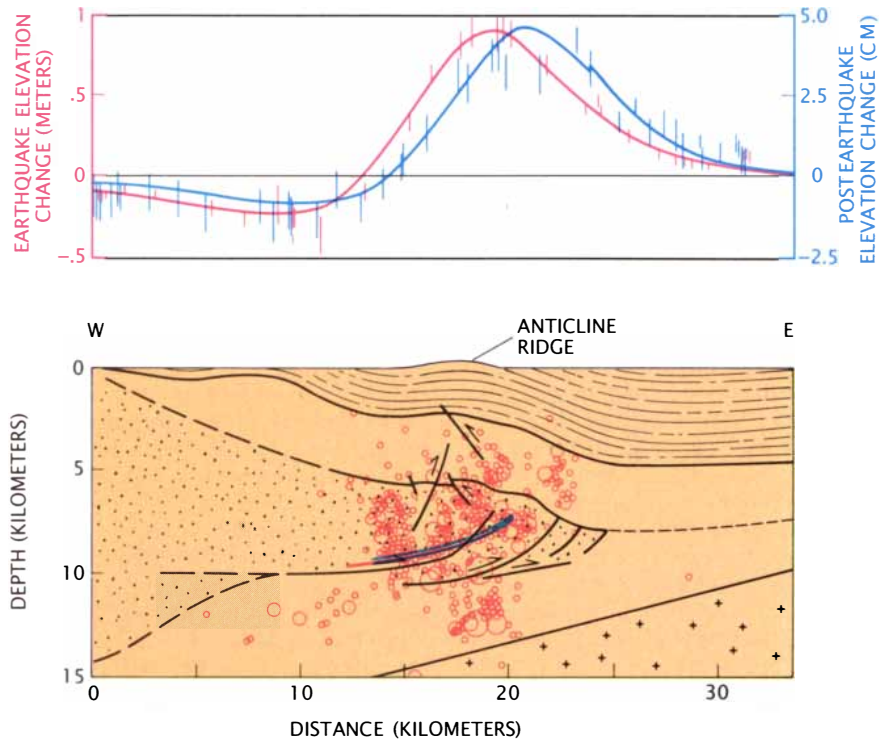
able and unexpected. The more important question, though, is whether such active folds conceal large faults, which in turn could provide the sites for large shocks.

By virtue of hindsight it appears that during the past half century large fold earthquakes have taken place in Japan, Argentina, New Zealand, Iran and Pakistan, giving an apparent affirmative answer to the question. The evidence for these events, however, is sparse and subject to interpretation. Yet since 1980 four major events have unequivocally established the reality of fold earthquakes. A fifth event may soon be added to the list.

The first of these shocks took place in 1980 in El Asnam, Algeria. It measured 7.3 on the moment-magnitude scale (a successor to the Richter scale, designated **M**), killed 3,500 people in three North African cities and dammed a major river. The second and third earthquakes, the 1983 **M** = 6.5 Coalinga shock and the adjacent 1985 **M** = 6.1 Kettleman Hills event, shook areas of California. Because these sites are remote, the earthquakes caused only modest damage and one death, but the areas did include several major toxic-waste storage sites, and in the small town of Coalinga, 75 percent of the unreinforced buildings were demolished. The next shock struck within California's populous Los Angeles basin on October 1, 1987. Although this **M** = 6.0 Whittier Narrows quake was only one tenth the size of the Coalinga event, it caused 10 times the damage—$350 million—and took eight lives. The fifth, candidate event was the tragic 1988 Armenian earthquake, which claimed at least 25,000 lives.

The most obvious feature of the three California earthquakes was that none of the faults that slipped cut the surface of the earth. In the Algerian event the fault did cut the earth's surface, but the amount of fault slip at ground level was far less than the amount of slip at the depth of the earthquake's focus, about 10 kilometers down. These observations suggest that in fold earthquakes the slip diminishes from earthquake level toward ground level, and so there is little, if any, rupture of the surface.

A second feature was that all four earthquakes took place under young anticlines—anticlines that are less than several million years old. Taken with the third and most revealing observation—that at each site the fold rose perceptibly during the earthquake—this suggests not only that young anticlines mark potential earthquake sites but that folds are actually the geologic product of successive earthquakes. If the pattern of diminishing fault slip toward the surface holds in general, then folds should enable us to divine the history of the hidden, or "blind," faults below.



COALINGA EARTHQUAKE of M = 6.5 in 1983 took California by surprise because it was not associated with any surface fault. The deformation of the overlying fold (plotted against distance from an arbitrary origin) is shown by the red vertical lines in the top panel. The red curve represents the deformation produced by a model blind fault, shown in red in the bottom panel. Deformation in the four years since the event is shown in blue in the top panel. The ongoing deformation indicates that the fault continues to propagate into the fold's core, as shown by the blue curve in the bottom panel. Seismic profiling reveals many faults deep in that region of the anticline (*short lines, bottom panel*). Small displacements along them during the 1983 event may explain the widely distributed pattern of aftershocks (*circles*) that followed the main shock (*largest circle*). Such a pattern contrasts strongly with that of a surface-fault earthquake, in which all the aftershocks are aligned with the fault.

Unsurpassed data from seismology, geology and geodesy—the measurement of the shape of the earth's surface—make the Coalinga event the most vivid illustration of a fold-related earthquake. Anticline Ridge, a 750-meter-high fold (appropriately named by the petroleum geologists who discovered oil there in 1898), grew 75 centimeters during the earthquake [*see illustration above*]. The youngest folded sediments are less than two million years old, which indicates that the fold began to form since that time. If a 75-centimeter growth is typical for events at the Coalinga site, then Anticline Ridge could have been built by roughly 1,000 ancestral earthquakes of magnitudes from **M** = 6 to **M** = 7 recurring every 1,000 to 2,000 years.

The Coalinga event also differed from surface-fault earthquakes in the pattern of its aftershocks. Typically the aftershocks of surface-fault earthquakes are aligned along the plane of the fault itself. On the other hand, the aftershocks at Coalinga—and of fold earthquakes in general—are distributed much more diffusely, both above and below the fault plane. What does this tell us?

Carl M. Wentworth, Jr., of the U.S. Geological Survey and Mark D. Zoback of Stanford University have investigated the deep geologic structure at Coalinga with the help of data from oil-well drill holes, seismic-refraction profiling (based on the travel time of sound waves excited by a controlled explosion as they traverse the structure) and seismic-reflection profiling (based on the timing of sound waves that are generated by orchestrated

thumping of the ground and reflected off buried strata). Their results suggest that the core of the Coalinga anticline is riddled with faults, many of which exhibit displacements that make up a small fraction of the displacement along the major fault lying at a greater depth.

The majority of structures imaged by these techniques are "reverse" faults, steep faults along which the younger strata are pushed over the older beds below, like a carpet pushed up a tilted floor. When reverse faults are less steeply inclined, they take on the name "thrust" faults. The main Coalinga shock appears to have nucleated on one of the major blind thrust faults. Many of the aftershocks, however, appear to be the result of slip on the smaller reverse faults. The pattern stands in strong contrast to that of a surface-fault earthquake, in which the aftershocks generally take place along a single plane (as can be seen along the San Andreas fault in the illustration on pages 48 and 49).

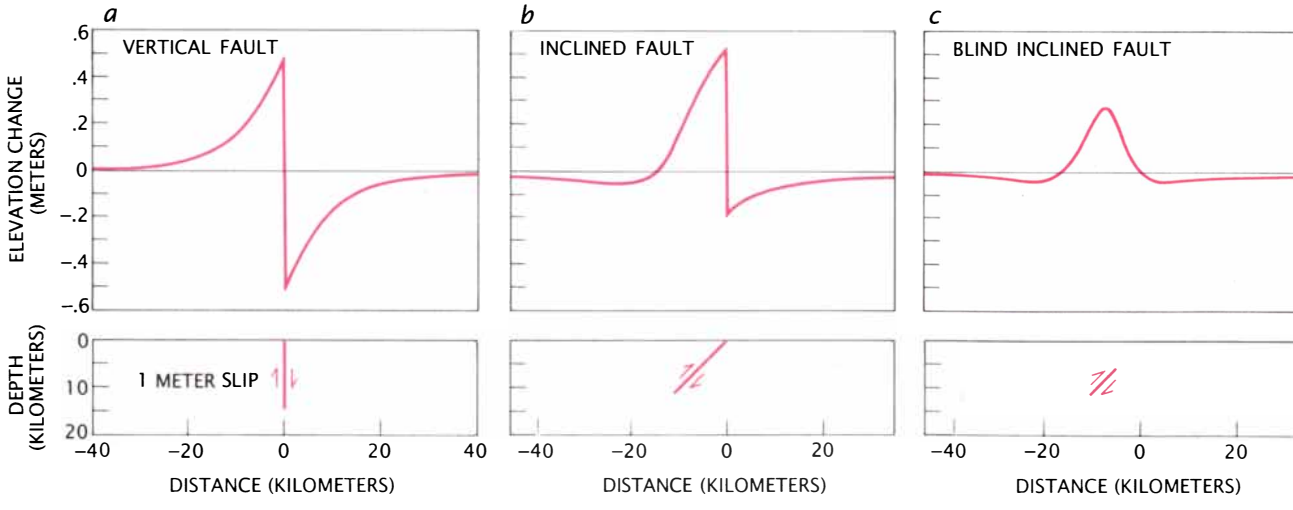
How does one measure the displacement of faults hidden beneath the surface? The location and amount of slip cannot in fact be observed directly, but they can be deduced from surface measurements with the help of a simple elastic model of the earth's crust. One treats the crust as an elastic solid with an embedded cut—similar to a stiff block of rubber that has been sliced with a knife. When the two faces of the cut are slid past each other, the rubber is strained and its surface is deformed. This elastic analysis was first developed to understand angstrom-sized flaws in crystals but has since been extended to study ruptures of faults 1,000 kilometers long.

By such methods we have discovered that the fault displacement at Coalinga, which was about 3.5 meters at the earthquake depth of nine kilometers, diminished upward, giving rise to folding at shallower depths, and disappeared at a depth of about six kilometers. After the earthquake the fault continued to slip: over the next four years the surface fold deformed by another 10 percent, and the axis of uplift shifted several kilometers eastward [*see illustration on preceding page*]. The simplest explanation for this observation is that the fault tip is propagating upward and eastward into the core of the anticline. The fault migration is due to the fact that the stress imparted by the earthquake is greatest just beyond the tip of the fault. Any visitor to Independence Mall in Philadelphia has observed the phenomenon: curators of the Liberty Bell were compelled to drill a cylindrical hole at the end of its famous crack to distribute the stress evenly and prevent this national treasure from splitting in two.

Given the tendency of fault tips to propagate, one might well wonder why the great majority of thrust faults, such as the one beneath Coalinga's Anticline Ridge, do not extend all the way to the surface. Through analysis of oil-well logs John Suppe of Princeton University has identified active folds in Taiwan that result from slip along deeply buried faults. He has also examined ancient folds, in which the overlying sediments have been exposed, thereby exhuming the faults beneath. The studies reveal that these faults form at depths of 10 to 20 kilometers and propagate slowly toward the surface over millions of years. Consequently many faults are undoubtedly still propagating to the surface and have not yet signaled their presence. Other, inactive faults will remain invisible until they are exposed by erosion of the overlying rocks. This pattern contrasts with that of the more familiar faults, where "what you see is what you get"; such faults either propagate to the surface rapidly or propagate downward.

By modeling the fault-propagation process, Suppe has also found that as the faults migrate, so do the fault-propagation folds, which grow in amplitude and change shape [*see illustration on opposite page*]. The postearthquake deformation at Coalinga shown on the preceding page supplies a remarkable snapshot of the process envisaged by Suppe: as the fault tip propagated to the right, the fold also migrated in that direction.

One example of a fold-related earthquake would not constitute a class. Yet the Coalinga event was not an isolated case; in fact, it was the second of a trio of related events that took place along the same blind thrust fault, the first being the



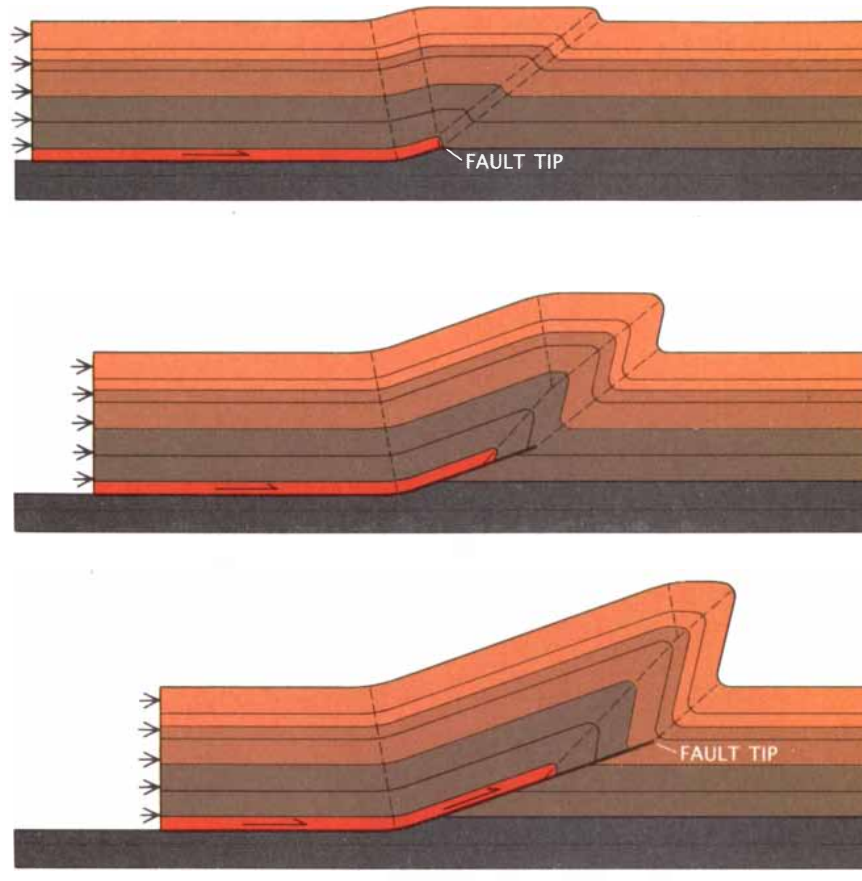
ELASTIC MODEL demonstrates how a blind fault deep underground (*bottom panels*) can deform the earth's surface (*top panels*). The faults are cut into the earth to a depth of 15 kilometers, much like slices that have been made by a knife in a rubber block. The left side of the block moves upward one meter with respect to the right side. If the fault is vertical (*a*) and extends to the surface of the earth, the surface deformation is one meter at the fault axis and asymptotically approaches zero far from the fault. If the fault is inclined 45 degrees (*b*), the total surface displacement is only 70 percent (cosine 45 degrees) as large. If the fault does not reach ground level (*c*), the fault is termed blind, and the slip results in a fold at the surface.

October 25, 1982, New Idria earthquake to the north and the third being the August 4, 1985, Kettleman Hills earthquake to the south. The three earthquakes were sudden responses to northeast-southwest compression along what is evidently one regional thrust fault approximately 100 kilometers long. The fact that the earthquakes proceeded from north to south suggests that a slip on one patch loads an adjacent patch to the south, which ruptures next. Although the south half of this string of anticlines has not produced a large historical earthquake, it must be viewed as a candidate for a future shock. Repeated earthquakes along the blind fault must have taken place to build this long chain of folds.

Stronger evidence for this point of view comes from the 1980 **M** = 7.3 El Asnam earthquake in Algeria. The earthquake was produced by a three- to six-meter slip on a reverse fault several kilometers underground. Only in the central segment of the fault did much of the slip (two meters) reach the surface. Geoffrey C. P. King, now of the U.S. Geological Survey, Claudio Vita-Finzi of University College, London, and J. C. Ruegg and his colleagues at the Institute of Physics of the Globe in Paris showed that an anticline associated with the fault was uplifted five meters during the earthquake and that the adjacent valley buckled downward one meter [*see top illustration on next page*]. As a result the Chéliff River, which flowed through a gorge it had hewn through the growing anticline, was dammed. Within days after the earthquake the Chéliff River swelled into a lake.

Before the river cut through the fold 40 centimeters of silt had accumulated in the lake, leaving a permanent mark in the geologic record. Mustapha Meghraoui of the University of Paris and his colleagues have excavated the underlying deposits and found that over the past 6,000 years there has been a succession of six such short-lived lakes, each marking the onset of sudden flooding. The best explanation is that these lakes were the products of surface-folding earthquakes. The El Asnam site not only furnishes evidence for a single fold earthquake but, more important, contains an archive of fold earthquakes, once again providing evidence that anticlines are built by a series of discrete events.

El Asnam displays further evidence that it has experienced repeated earthquakes. Small, secondary faults on the top of folds can also slip during an earthquake in response to the buckling of the fold. If they do, the slips form an accessible record of past quakes, which can be dated by applying carbon-14 methods to the surrounding sediment. Such dating was carried out by Meghraoui and his colleagues on secondary faults in the El Asnam region, and the results are consistent with the lake-sediment record.



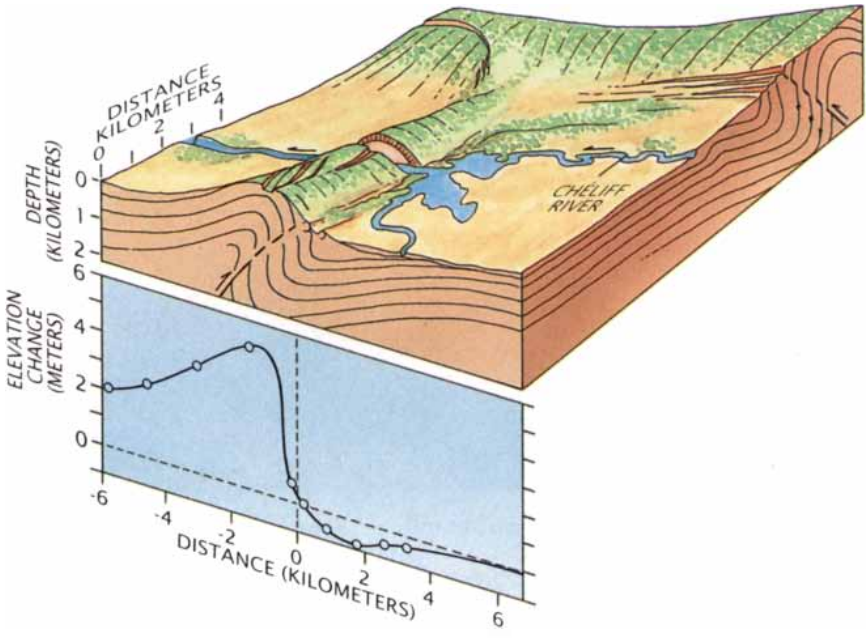
FAULT-PROPAGATION FOLD was modeled by John Suppe of Princeton University. The rock slips along a horizontal fault in a weak stratum and deforms (*top*). The fault then splays upward and propagates into the overlying beds (*middle*). Continued displacement of the rock along the fault causes the fold to grow upward, widen and change shape (*bottom*). Such a process may be active in the Coalinga fold (*see illustration on page 51*). Several effects that influence the geometry of folds, including the buoyancy and elasticity of the crust, sedimentation and erosion, have been neglected.

The small but ominous Whittier Narrows event, which took place in 1987 in the Los Angeles metropolitan area, rounds out the recent evidence for fold earthquakes. Like its larger predecessors, this earthquake appears to be the product of a tear in a stack of sediments that are being compressed in a geologic vise between two moving crustal plates. By graphically unfolding the buckled strata in the Los Angeles basin, Thomas L. Davis and Jay Namson of Davis and Namson Geological Consultants in Los Angeles have argued that the basin has been shortening at the rate of about a centimeter per year for the past 2.2 million years, which represents 20 percent of the entire plate motion of the western U.S. concentrated in a zone 50 kilometers wide. In other words, Los Angeles may be losing one-quarter acre each year to active surface folding, which may explain the concentrated and enigmatic seismicity of the basin.

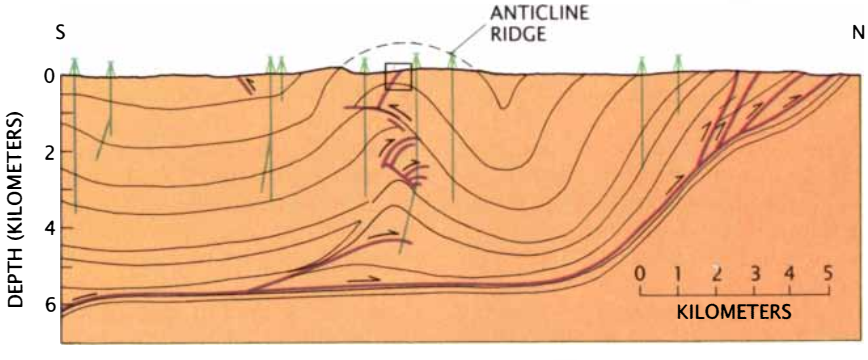
Like its predecessors the Whittier Narrows earthquake also lifted its associated anticline, the Santa Monica Mountains fold. And like the New Idria, Coalinga and Kettleman Hills shocks, the Whittier Narrows event seems to mark an extensive blind fault. Evidence from the alignment of small-to-moderate earthquakes (advanced by Egill Hauksson of the University of Southern California and Lucile M. Jones of the U.S. Geological Survey) suggests that the Whittier Narrows event is but one of several earthquakes that have taken place along a blind fault running for 150 kilometers beneath the California

coastline. The **M** = 5.6 earthquake on February 21, 1973, at Point Mugu took place at the western end of the proposed fault and shares many attributes with the larger fold-related earthquakes: it accommodated crustal compression, uplifted the Santa Monica Mountains by 35 millimeters, left a diffuse zone of aftershocks and showed no surface-fault slip.

In the preceding discussion much of the evidence for fold earthquakes may have seemed indirect: projection of surface deformation to greater depths and examination of aftershock distribution. More direct evidence can be obtained only with access to the interior of an anticline, and direct access is possible only when a fold has been extensively plumbed for oil. The Ventura Avenue anticline, which is punctured by more than 1,400 wells drilled to a depth of as much as 6.6 kilometers, is among the best studied—and also fastest-rising—folds in the world and so is ideal for close inspection.



EL ASNAM EARTHQUAKE of M = 7.3 in 1980 dammed Algeria's Chéliff River. The river had cut a gorge through the fold at the earthquake site, but during the event the ground rose five meters, blocking the flow. Silt accumulated in the transient lake that was created thereby; excavations under the new silt layer revealed evidence for six such short-lived lakes over the past 6,000 years. They are likely to have been the product of fold-related earthquakes that have recurred every millennium. The graph shows the change in elevation of the ground across the fold; the uplift pattern's similarity to the shape of the fold suggests that such events actually built the fold.



INTENSE FRACTURING pervades the core of the Ventura Avenue anticline. Such fracturing, caused by bending of the strata, produces many faults (*purple lines*). Analysis of data from oil wells (*vertical green lines*) drilled into the anticline shows that the fold flattens at a depth of about six kilometers. This is best explained by the presence of a nearly horizontal fault, or thrust fault, at the fold's base. Movement along faults that diverge from this main fault causes the fold to buckle and grow as does the model on the preceding page. The box indicates the region shown on page 50.

The Ventura Avenue anticline, located on the southern California coast, has uplifted former deep-water sediments and exposed these folded layers in cliffs near the coastline. Although no major earthquakes have been registered along the Ventura Avenue anticline, its structure and prehistoric record tell us much that is pertinent to the other earthquake sites we have discussed.

Like many anticlines the Ventura Avenue structure traps oil and gas that migrate into its core. The logs of the oil wells that have been drilled into the Ventura Avenue anticline enable us to view it in cross section [*see bottom illustration at left*]. The curvature of the strata is broad at the surface but becomes progressively tighter at greater depths. At about four kilometers below ground level, the fold exhibits a sharp kink, underlaid by mechanically weak rock. An exploratory well 6.6 kilometers deep penetrated enough of the weaker rock to demonstrate that the strata below the kink are not folded at all. This evidence suggests that beneath the fold lies a nearly horizontal thrust fault. Smaller, steeply inclined reverse faults extend upward from the base of the anticline to ground level.

The Ventura anticline appears to have been uplifted by slip on the Barnard thrust fault that branches off the Sisar fault at the base of the fold. The slip of the internal reverse faults also seems to result from motion along the Barnard fault. Farther to the north, slip on the Barnard fault exceeds that on the Sisar fault, causing the strata to be jammed together and bowed upward. This close inspection of the Ventura anticline shows that it not only hides a major thrust fault at the depth of the greatest slip but also harbors myriad smaller reverse faults within its core. These structural features are remarkably similar to those inferred from seismic imaging and geodetic measurements at the Coalinga and Kettleman Hills folds.

The internal structure of the Ventura Avenue anticline helps to explain several shared features of the Coalinga, Kettleman Hills, Whittier Narrows and El Asnam earthquakes: although the greatest slip takes place on the major fault at the fold base, slip on the abundant small

reverse faults scattered throughout the core may well account for the diffuse distribution of aftershocks that characterizes fold earthquakes.

Another observation from Ventura Avenue helps to explain a second distinguishing feature of most fold events, which we have not yet mentioned: the faults ruptured slowly. Although not all fold-related earthquakes are slow, almost all slow earthquakes are fold-related. Göran Ekström of Columbia University has shown that the Kettleman Hills rupture took about four times longer (16 seconds) than is typical for a surface-faulting earthquake of the same size. John L. Nábelêk of Oregon State University found that at the northern segment of the El Asnam site, where the aftershocks were most dispersed and negligible reverse faulting cut the surface, the seismic rupture took twice the expected time. At Coalinga the seismic rupture time was normal, but the fold continued to grow after the earthquake at an exponentially decaying rate, which was a result of continued slow slip and propagation of the blind thrust fault.

The slow rupture time may be explained by the high fluid pressures registered in boreholes at the Ventura Avenue, New Idria, Coalinga and Kettleman Hills sites. Robert F. Yerkes of the U.S. Geological Survey has shown that the fluid pressure there exceeds hydrostatic pressure below three kilometers. The folds seal in the fluids, whereas surface faults provide a source of leakage. High fluid pressure in the rock fractures reduces the frictional resistance of the rock to sliding, thereby enhancing its susceptibility to both faulting and folding. Diffusion of the pore fluid after the rupture alters the friction along faults, inhibiting slip on some fractures and promoting it on others. The result is a spread of rupture times that can manifest itself as a slow, or prolonged, earthquake.

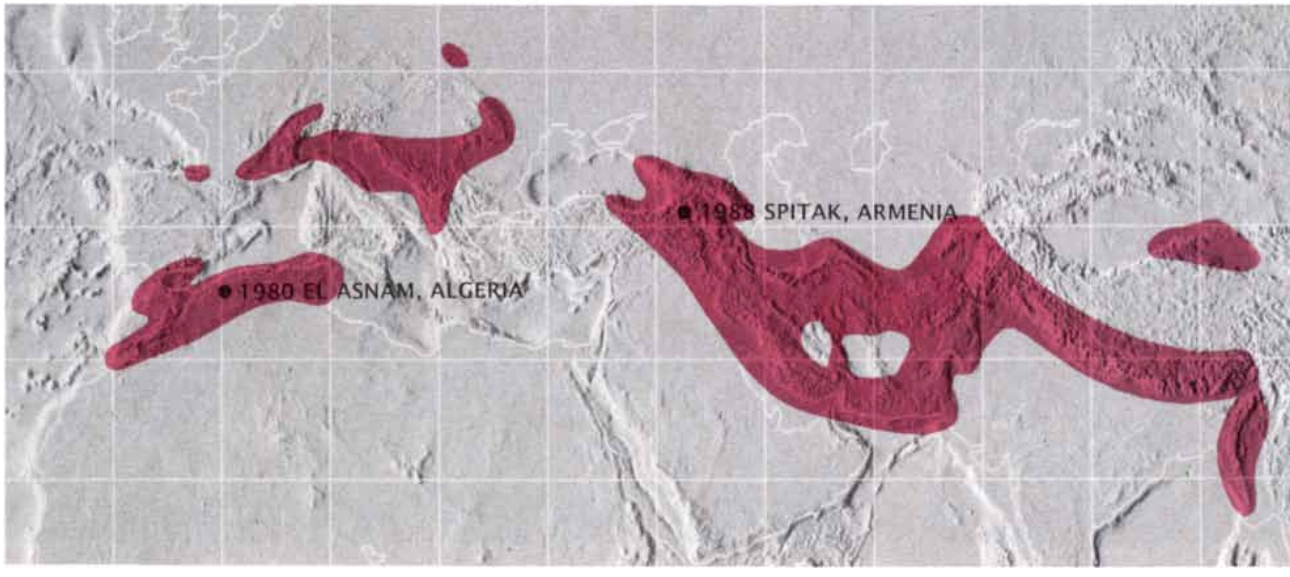
We claim that folds are built—at least in part—in a punctuated fashion by successive earthquakes over hundreds of thousands or millions of years, rather than by slow, steady deformation. The Ventura Avenue anticline also provides persuasive evidence for this point of view.

The anticline is only 200,000 to 300,000 years old; during that period it has been growing at one of the fastest rates in the world. Kenneth R. Lajoie and Andrei M. Sarna-Wojcicki of the U.S. Geological Survey have identified nine marine platforms on the flank of the anticline, the youngest one lifted two meters above sea level, the oldest raised 20 meters. Each of these terraces was cut by storm waves at or near sea level and strewn with seashells. Carbon dating of the shells yields ages of 1,800 years for the youngest terrace and 5,600 years for the oldest. Because sea level is known from independent evidence to have been within a meter of its current height during the past 7,000 years, the terraces could not have been cut during periods when sea level was higher than it is now; rather the land has been rising. The preservation of nine separate platforms suggests that this uplift was not gradual but sudden. As in the case of the El Asnam lakes, the simplest explanation for this giant flight of stairs is uplift produced by repeated earthquakes.

We have argued that the El Asnam, Coalinga, Kettleman Hills and Whittier Narrows earthquakes provide indisputable evidence that folds grow during large earthquakes. Moreover, the prehistoric record at El Asnam and Ventura Avenue also illustrates that such events have recurred. Yet most earth scientists have viewed folds as evidence for steady, progressive deformation that need not be punctuated by earthquakes. The difficult question, then, is: Are the cases we have discussed the exception or the rule? In other words, is fold growth dominantly steady and aseismic or do most actively growing folds in fact conceal seismic faults?

If one were to ask the same question about motion along active faults that cut the earth's surface, the answer would be that few faults are seen to undergo steady slip, or creep; the rule is that creep is uncommon in the uppermost 15 kilometers of the crust. There are of course notable exceptions, the central 300 kilometers of the San Andreas fault among them. Worldwide, however, most faults exhibit the stick-and-slip behavior that gives rise to earthquakes.

Nowhere is this question more important than in the Los Angeles basin. The 1.5-kilometer-high Santa Monica



ACTIVE FOLD BELTS girdle the earth. Most fold belts, the result of crustal compression where continental tectonic plates collide, form extensive mountainous regions. Much of the world's population lives within these zones, which include parts of Italy, Yugoslavia, Iran, Afghanistan, Pakistan, India, Nepal, Burma, Japan, the U.S., Peru and Chile. This map was compiled principally from the work of Albert W. Bally of Rice University and Hervé Philip of the University of Montpellier in France.

Mountains fold (which, at Whittier Narrows, is almost completely buried by sediments shed by the San Gabriel Mountains) is less than 3.3 million years old. The average uplift rate is therefore about .5 millimeter per year. During the Whittier Narrows earthquake, however, the fold grew 50 millimeters at the earth's surface. If fold growth is caused entirely by earthquakes, then an earthquake should recur at this site every century. Yet the section of fault that slipped during the 1987 earthquake was about five kilometers long, only about 3 percent of the length of the entire blind fault, which may be as long as 150 kilometers. Consequently, if all parts of the fault are equally likely to slip in an **M** = 6 earthquake, shocks should strike the Los Angeles basin and the adjacent coastline every three to five years. Clearly this has not happened: an average of just one shock of magnitude greater than or equal to **M** = 6 has been recorded every 25 years in this corridor since Los Angeles was settled in the early 19th century.

One possible explanation for the discrepancy is that most of the fault slip and fold growth is in fact steady: stress does not accumulate but is continuously relieved by creep. The thrust faults beneath many folds lie at depths of 12 to 15 kilometers, where the crust may reach a temperature at which the quartz component of the rock becomes ductile and facilitates creep. If the faults undergo steady slip, then the folds above may also deform progressively. Although there is no direct evidence for steady fold growth, stable sliding has been simulated in rock samples in the laboratory, and as much as half of the slip on the interface between tectonic plates along the plate boundaries around the Pacific Ocean is aseismic.

We think it is more likely that the folds grow intermittently through earthquakes that are larger than the Whittier Narrows event but that are much less frequent. Earthquakes of magnitudes greater than **M** = 7 recurring every several hundred years could account for the fold growth, and the scenario would be consistent with

ARMENIAN EARTHQUAKE SITE was mapped by *Landsat* in 1987 and later color-enhanced for the Soviet Mission in Washington, D.C. The main 1988 shock (*large white circle*) occurred near a limited surface fault (*heavy blue line*) not far from the town of Spitak. A string of folds (*red line*) mapped by Soviet geologists 25 years before the earthquake is seen to be aligned with the widespread aftershocks (*small white circles*), which were recorded by a U.S. National Academy of Sciences–U.S. Geo-

the absence of smaller events during the brief recorded history of the Los Angeles basin.

It is incumbent on seismologists to distinguish between the two competing explanations; their consequences are dramatically different. If earthquakes larger than the 1987 event are possible beneath the Santa Monica Mountains fold, then Los Angeles' greatest earthquake threat may come not from a future **M** = 8 shock on the San Andreas fault, 50 kilometers to the north, but from a smaller earthquake originating under downtown Los Angeles.

Several tools are now at our disposal to probe the Los Angeles basin as well as other active fold belts around the world. They include seismic-reflection profiling, geodetic measurement of the present rate of fold contraction and uplift, examination of the logs from thousands of oil wells and geologic investigations to identify the fastest-growing surface folds.

Although southern California has served as a focal point for efforts to study fold-related earthquakes in the U.S., evidence indicates that such events have struck throughout the world's active fold belts and therefore endanger a much larger population [*see illustration on page 55*]. In addition to North Africa, the site of the El Asnam earthquake, events of magnitudes from **M** = 7 to **M** = 7.8 have taken place on blind faults in northern India, New Zealand, Argentina, Canada and Japan. Countries such as Chile, Yugoslavia, Taiwan, Iran and Pakistan may also be subject to highly damaging fold events.

The most recent—and among the most tragic—example of a fold earthquake claimed worldwide attention just six months ago. We are of course referring to the December 7, 1988, **M** = 6.8 earthquake in Spitak, Armenia, which killed at least 25,000 people. The earthquake struck in one of the world's most intensely folded and seismically active regions, the Lesser Caucasus Mountains of the Soviet Union. The main shock was the result of slip on a reverse fault beneath a youthful anticline. Rupture along a surface fault was found, but it extended for only eight to 12 kilometers, whereas aftershocks were distributed over a 50-kilometer-long zone beneath the fold. This suggests that little of the slip that caused the earthquake reached the earth's surface. As with the other fold earthquakes we have discussed, the aftershock pattern was also diffuse rather than aligned along the fault plane, as it would have been in a surface-fault earthquake. The preliminary evidence therefore indicates that the Spitak event was fold-related.

The size and destructive character of the Armenian earthquake could not have been predicted on the basis of its short surface fault, but the fold provided a clue that seismologists must not, in the future, overlook.

logical Survey seismic team at the invitation of the Academy of Sciences of the U.S.S.R. The limited surface rupture and the close association of the fold and aftershocks suggest that the earthquake was fold-related. It is not clear whether the major but heretofore unverified fault (*thin blue line*) slipped during the event.

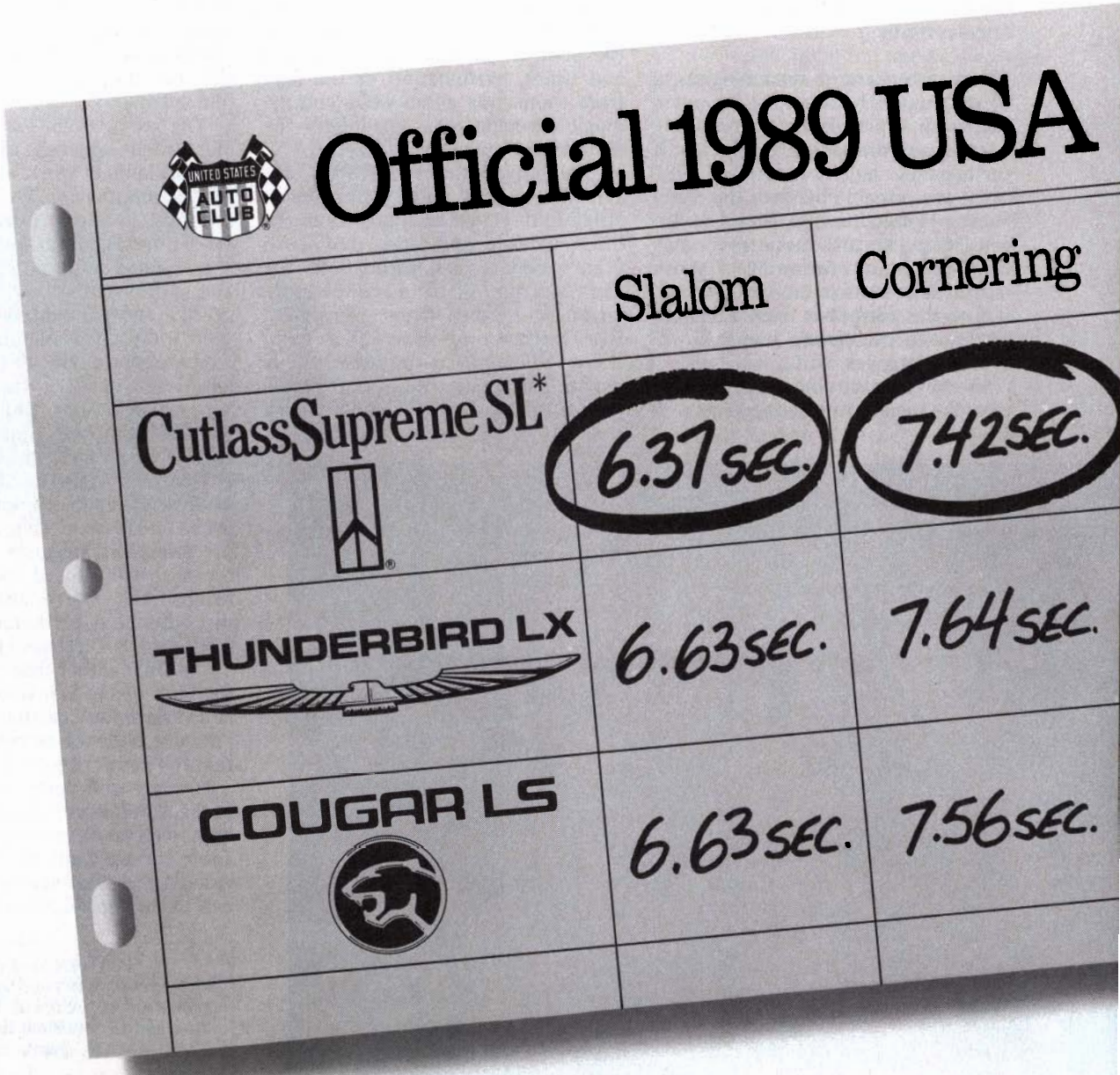
FURTHER READING

ACTIVE FOLDING IN THE ALGERIAN EARTHQUAKE OF 10 OCTOBER 1980. G. C. P. King and C. Vita-Finzi in *Nature,* Vol. 292, No. 5818, pages 22-26; July 2, 1981.

SEISMIC POTENTIAL REVEALED BY SURFACE FOLDING: 1983 COALINGA, CALIFORNIA, EARTHQUAKE. Ross S. Stein and Geoffrey C. P. King in *Science,* Vol. 224, No. 4651, pages 869-872; May 25, 1984.

ACTIVE FAULTS RELATED TO FOLDING. Robert S. Yeats in *Active Tectonics.* Panel chaired by Robert E. Wallace. National Academy Press, 1986.

THE 1987 WHITTIER NARROWS EARTHQUAKE IN THE LOS ANGELES METROPOLITAN AREA, CALIFORNIA. Egill Hauksson, Lucile M. Jones, Thomas L. Davis, L. Katherine Hutton, A. Gerald Brady, Paul A. Reasenberg, Andrew J. Michael, Robert F. Yerkes, Patrick Williams, Glen Reagor, Carl W. Stover, Allison L. Bent, Anthony K. Shakal, Edwin Etheredge, Ronald L. Porcella, Charles G. Bufe, Malcolm J. S. Johnston and Edward Cranswick in *Science,* Vol. 239, No. 4846, pages 1409-1412; March 18, 1988.



Oldsmobile: five.

In the sophisticated world in which the new Oldsmobile Cutlass Supreme competes, it would be presumptuous even to consider singling out any one car as "The Best."

So when a Cutlass Supreme SL, equipped with the optional FE3 suspension, was pitted against the new Ford Thunderbird LX and Mercury Cougar LS in five recent USAC tests,* it was logical to assume, as the saying goes, "They'd win some, and we'd win some."

They didn't win any.

Borne off the line by a new 3.1-liter V6, Cutlass Supreme outpowered them from zero to 60.

It beat them handily through the slalom course, its optional FE3 suspension taking the pylons in textbook form.

In a test of passing power, it readily outaccelerated them from 40 to 65 mph.

On the skidpad, it stuck like flypaper, pulling 0.83 g's to easily best their lateral acceleration capabilities.

And in a test of braking, it stopped 13 feet shorter than Thunderbird—16 feet shorter than Cougar.

Rarely does a battery of tests come out this overwhelmingly in favor of one car. A fact that, in itself, probably says something about the rarity of a car like the new Cutlass Supreme.

And no matter how you grade a car's

Let's get it together ...buckle up.

GM

C Test Results

UNITED STATES AUTO CLUB

Braking	40-65 Passing	0-60 Acceleration
148.74 FT.	6.44 SEC.	9.36 SEC.
162.32 FT.	6.87 SEC.	10.46 SEC.
165.49 FT.	6.81 SEC.	10.33 SEC.

*with optional FE3 suspension.

Ford: zip.

performance—whether by stopwatch and lateral g-forces like USAC, or when passing a semi and winding through the mountains like the rest of us—we feel that *this* car will do well every time. It even won a test USAC didn't conduct. In complete car coverage, its 3-year/50,000-mile Bumper to Bumper Plus warranty** can't be beat. And it's priced $1,200 less than Thunderbird LX—$200 less than Cougar LS.†

For more information about the car that swept the USAC tests, call toll-free, 1-800-242-OLDS, 9 a.m. to 7 p.m., EST. Better still, stop by an Olds dealer where you can arrange a test drive of your own. We think that even if you don't take it to the skidpad, you'll be impressed.

The New Generation of OLDSMOBILE

*USAC tests based on comparisons of Cutlass Supreme SL with optional FE3 suspension and Ford Thunderbird LX and Mercury Cougar LS.

**See your Oldsmobile dealer for terms of this limited warranty.

†Based on MSRP's of cars as tested by USAC. Levels of equipment vary.

A Different Kind of Inheritance

The methylation of DNA may be a major "epigenetic" mechanism by which gene-activity patterns—as opposed to genes per se—are passed from one generation of cells to another during development

by Robin Holliday

New parents are not alone in wondering how the union of a single sperm and egg can lead to the birth of a fully formed individual, complete with functionally distinct tissues and organs. The puzzle is also of major concern to biologists.

Clearly part of the answer resides in the genes of the new embryo, which are inherited from the parents in the chromosomal DNA of the sperm and egg. As genes are switched on or off (permanently or transiently) in a predetermined sequence, they alter the mixtures of proteins in the newly forming cells, thereby altering the cells' activities. The earliest embryonic cells appear to be unspecialized, even though they have the potential to give rise to a complex organism. Later generations of cells become determined (committed to particular avenues of development); they then differentiate to become the specialized units of mature organisms. Cells in the differentiated state synthesize the so-called luxury proteins that give distinct tissues their unique identities.

The presence of a developmental program does not explain why early embryonic cells later diverge. Some process must act on the DNA to direct the unfolding of the program: to select particular genes to be switched on or off in particular cells at particular times. Otherwise all the somatic (nongerm) cells in the body, which have the same chromosomes, would develop identically. Something must also control the transmission of on-or-off patterns from one generation of cells to the next. This transmission is crucial if daughter cells are to retain or advance the specializations of their parents.

The processes that control the unfolding of the developmental program are described as epigenetic. The epigenetic rules for the alteration and inheritance of gene activities in the cells of the embryo, or in the more specialized somatic cells of the mature organism, are complex and are still far from being understood. This contrasts with biologists' extensive knowledge of classical genetics—the transmission of genes per se from parents to child. Nevertheless, the young field of epigenetics is now beginning to make some progress.

Most of the investigators who are trying to understand the control of gene activity in development would agree that it must be based on critical interactions between DNA and proteins. (After all, the binding of specific proteins to particular DNA segments in the regulatory regions of genes apparently controls gene expression in general: the transcription of a gene into messenger RNA by an enzyme known as an RNA polymerase and the subsequent translation of the messenger RNA into the encoded protein.) Beyond that there is much disagreement.

Proponents of one view, which is sometimes known as protein inheritance, hold that the complex formed by DNA and its bound proteins determines the heritable characteristics of a cell. They believe not only that the complexed proteins influence which genes are switched on and off during development but also that DNA-protein complexes are preserved in successive cell generations—that specific proteins remain tightly bound to particular regions of DNA right through chromosome replication and cell division. Just how such continued binding would be achieved has not yet been explained.

A significantly different possibility is that the genetic sequences themselves might be altered in specific ways and that those alterations would then be maintained by the normal processes that ensure the fidelity of DNA replication. It is clear, for example, that parts of the genes encoding antibodies are indeed rearranged in the course of differentiation, thereby ensuring that different immune system cells produce distinct antibodies. The rearrangements are not reversible, however, whereas there is evidence that epigenetic changes in the DNA of certain specialized somatic cells can be reversed or reprogrammed. Hence, DNA rearrangement probably is not a major mechanism directing development.

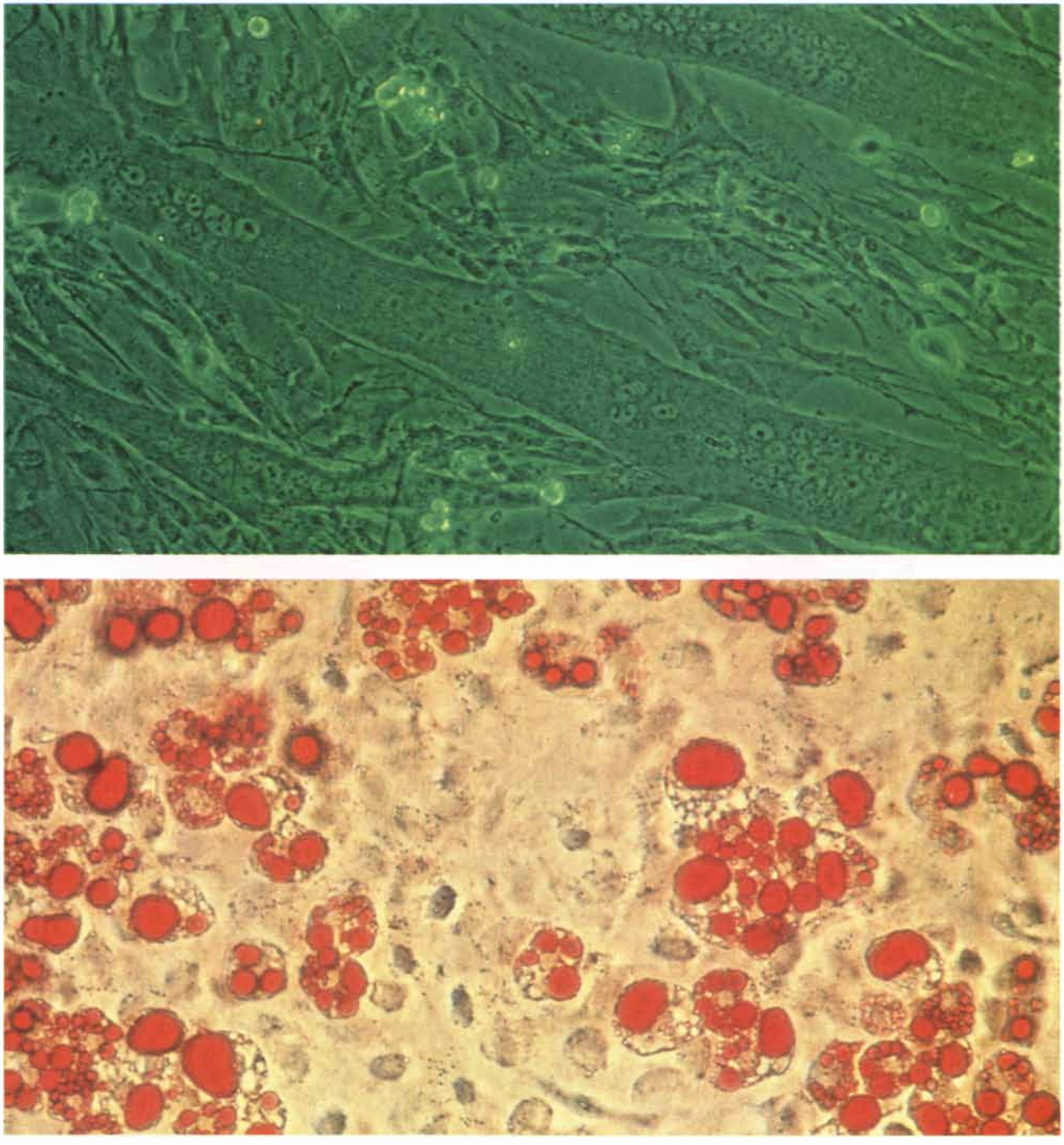
For some time a number of us have put forth a third idea, which is related to but distinct from the notion of protein inheritance. We think that the chemical modification of DNA may be an important epigenetic process in higher organisms. In particular, the attachment of a methyl group (CH_3) to cytosine, one of the four nitrogenous bases that distinguish the nucleotide building blocks of DNA, may participate in directing the genetic switching in developing cells. (The other bases are guanine, adenine and thymine.) At the least, cytosine methylation seems to be involved in controlling the cell-to-cell inheritance of gene-activity patterns. We suggest that methyl groups on DNA, rather than particular proteins, are transmitted intact from generation to generation of cells. We also suggest that the presence or absence of methyl groups acts as a signal to regulatory proteins, so that the pro-

ROBIN HOLLIDAY is a research scientist at the CSIRO Laboratory for Molecular Biology in Sydney, Australia. After earning his Ph.D. from the University of Cambridge in 1959, he joined the staff of the John Innes Institute in Hertford, England, where he developed a theory of genetic recombination now known as the Holliday model. He moved to the National Institute for Medical Research in London in 1965, becoming head of a new genetics division in 1970. There he studied recombination, DNA repair and cellular aging and became interested in the role of DNA methylation as a mediator of the epigenetic control of gene expression. Holliday went to CSIRO in 1988 and now concentrates on the study of DNA methylation.

teins interact with the DNA of daughter cells in the same ways as they did with the DNA of the parents.

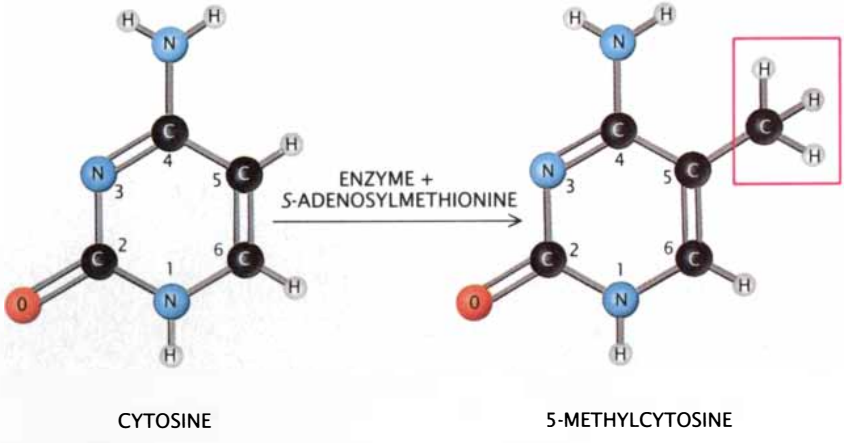
Various mechanisms by which DNA methylation could possibly help to control development were proposed before there was any evidence for or against them. Let me re-create history to some extent in describing the reasoning behind these proposals before I present evidence that has recently been discovered in their favor.

Interest in cytosine methylation of DNA as an epigenetic mechanism in eukaryotes was actually sparked by studies of the interactions of enzymes and DNA in bacteria, which do not have complex body plans. By early in the 1970's workers had found that so-called restriction enzymes recognized specific short sequences of bas-

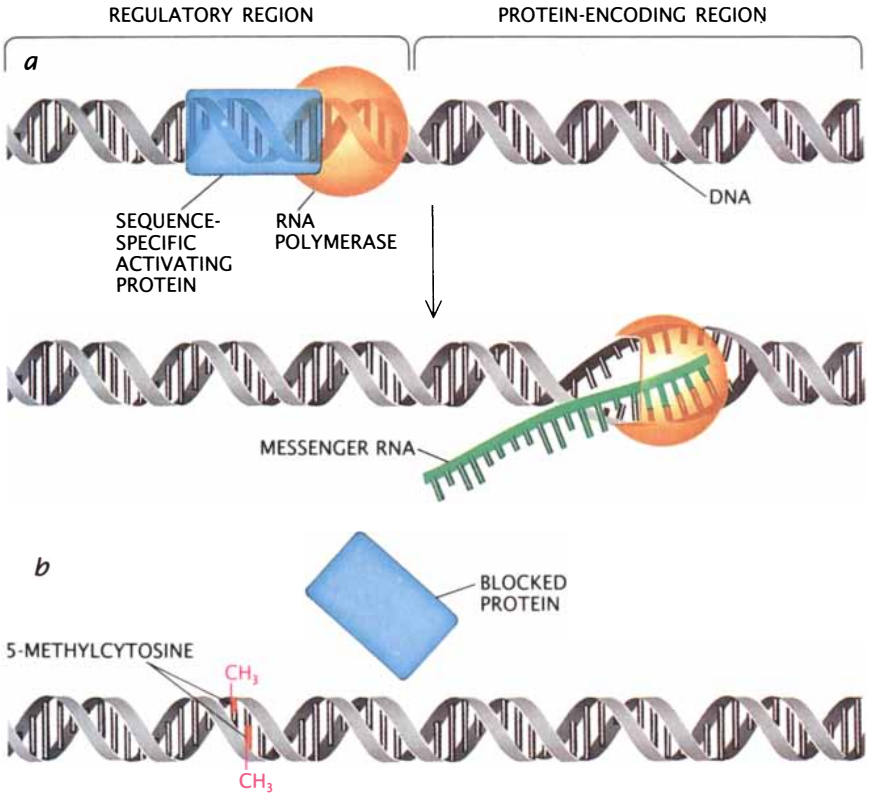


SPECIALIZED CELLS emerged when immortalized fibroblasts (connective-tissue cells) from mice were grown in the presence of azacytidine, an agent that removes methyl groups from DNA. Such demethylation is often associated with the activation of formerly silent genes. Normally fibroblasts do not change character, but in this instance some of the fibroblasts became muscle cells (*top*), which merged to form long fibers (*bands*) that could be seen to twitch. Other cells differentiated into adipocytes, or fat-storing cells (*bottom, stained red*). The findings provide evidence that alterations in the methylation status of DNA can influence the developmental pathways taken by cells. The photomicrographs were provided by Shirley M. Taylor, Lesley A. Michalowsky and Peter A. Jones of the University of Southern California School of Medicine.

es in DNA and cut the DNA at those sites. These enzymes were able to distinguish between sequences that were and were not modified in some way, and they cut only unmodified sequences. Among the most common modifications seeming to influence restriction-enzyme activity were the methylation of adenine and of what by convention is called the fifth carbon atom of cytosine, to form 5-methylcytosine. The methylation reactions were carried out soon after DNA replication by enzymes that had the same sequence specificity as the corresponding restriction enzymes.



CYTOSINE, one of four nitrogenous bases in DNA (*left*), can be methylated by an enzyme that replaces the hydrogen atom on the number-five carbon with a methyl group (CH_3) from the chemical *S*-adenosylmethionine to form 5-methylcytosine (*right*). Usually the modification of cytosine occurs soon after DNA has replicated.



TRANSCRIPTION of a gene (*a*) can be initiated when a protein (or proteins) recognizing a specific base sequence in the gene's regulatory region binds to the DNA; such binding enables an enzyme known as an RNA polymerase to transcribe the protein-encoding region of the DNA into messenger RNA. (The messenger RNA is later translated into the specified protein.) Methylation of the DNA (*b*), shown in red, can inactivate a gene by preventing the binding of an essential activating protein. Alternatively methylation might facilitate binding of an inhibitory protein (not shown).

The discovery that 5-methylcytosine could strongly affect the interaction of proteins and DNA, together with the well-established fact that 5-methylcytosine is a common constituent of DNA in vertebrates, plants and many other higher organisms, suggested to me and my student John E. Pugh (who were then working at the National Institute for Medical Research in London), and independently to Arthur D. Riggs of the City of Hope National Medical Center in Duarte, Calif., that the presence of 5-methylcytosine in DNA might influence gene activity in eukaryotes. It might do so by acting as a signal that facilitated either the activation or inhibition of genes. As I shall discuss in more detail below, a growing body of evidence now shows that methylation of the regulatory regions of genes (or nearby regions) is in fact generally associated with gene inactivation, presumably because the formation of 5-methylcytosine blocks the binding of transcription activators or else abets the binding of inhibitors.

Pugh and I, as well as Riggs, further proposed that the methylation signal might be retained by eukaryotic chromosomes as they replicate in preparation for mitotic cell division. All somatic cells are diploid, having a set of chromosomes from the mother and a paired set from the father. Before the cells divide, they duplicate their chromosomes: the two strands of each chromosome's DNA come apart, and new complementary strands are synthesized. Then a complete set of chromosomes is transmitted to each daughter cell.

Our proposed mechanism for inheritance was based in part on the knowledge that the sites often methylated in eukaryotic DNA are cytosine-guanine doublets: a cytosine followed on the same strand by a guanine. The bases on the paired strands of DNA are complementary, not identical; because cytosine on one strand of DNA always pairs with guanine on the complement, a cytosine-guanine doublet on one strand is always paired with a guanine-cytosine unit on the other strand. The configuration provides a substrate (cytosine) for methylation in almost precisely the same spot on both strands.

We postulated that an enzyme known as a maintenance methylase would act only on "hemimethylated" stretches of DNA, in which a cytosine-guanine doublet on one strand was

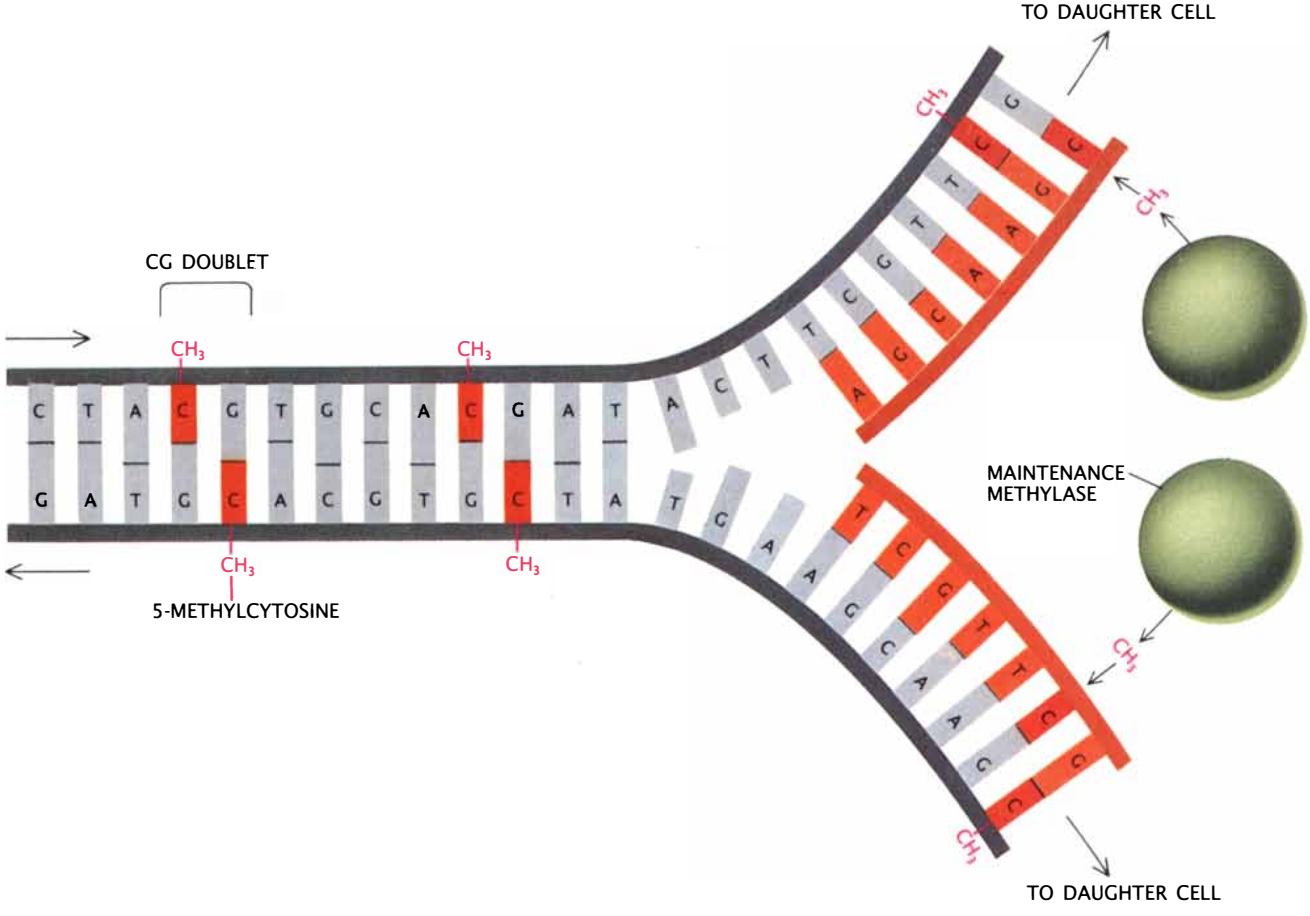
methylated but the complementary doublet on the other strand was not. When fully methylated DNA duplicates, it gives rise initially to hemimethylated DNA, which the maintenance enzyme would quickly methylate. In contrast, the progeny of unmethylated DNA would initially be unmethylated and would remain so because the maintenance methylase would not act on it at all. Hence, an existing methylation signal would be retained through cell division; the absence of methylation would also be maintained, and there would be no permanently hemimethylated regions. Under these rules the pattern of methylation (and nonmethylation) of DNA in a parent cell, and therefore a particular set of gene activities, would be re-created in the daughter cells.

This proposed scheme for maintaining the methylation pattern of DNA through many generations of cells does not explain how genes can be switched on or off during development, that is, how the protein-binding pattern is changed so that some cells become more specialized than their parents. No single model could account for all the switching in gene activities, but Pugh and I did suggest some ways in which the methylation or the demethylation of chromosomes could participate in such switching.

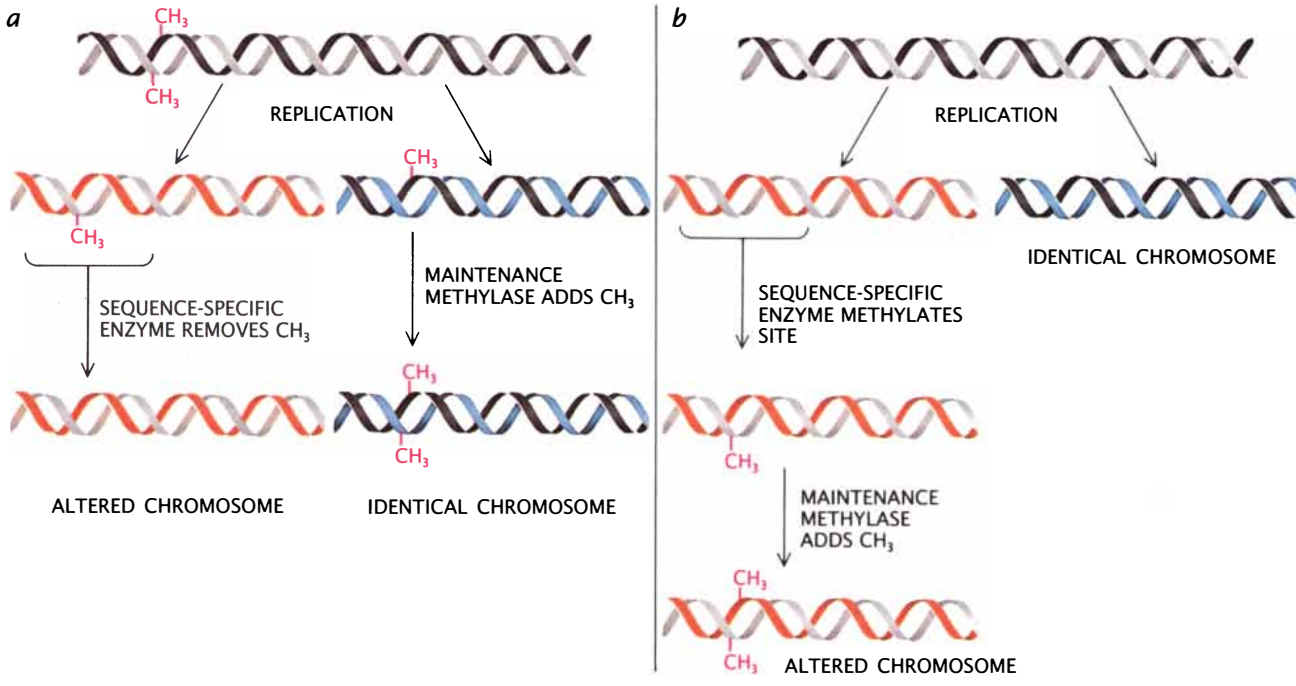
Our first assumption was that only a minority of cytosine-guanine doublets in DNA are involved in switching; DNA includes too many methylated doublets for them all to have a regulatory role. We also assumed that the doublets important in switching must be embedded in larger base sequences, which regulatory proteins specific to particular genes would recognize; alone, the ubiquitous doublets would provide no such specificity.

Given these assumptions, it seemed reasonable to suggest that switching might sometimes occur when an enzyme other than the maintenance methylase recognizes a specific base sequence in one strand of DNA in a chromosome and attaches a methyl group to a cytosine-guanine doublet within that sequence. This action would spur the maintenance enzyme to methylate the complementary strand quickly. Thereafter, future generations of the same chromosome would be methylated at the same site, even though methyl groups had not been attached to the gene initially.

In other instances, a sequence-specific enzyme might remove a methyl group from a particular site prior to cell division. In the absence of that methylation signal, the maintenance enzyme would not methylate the opposite site in the newly made complementary strand, and the formerly methylated gene would now be unmethylated. A different scenario would have a sequence-specific protein block maintenance methylation by binding to the spot where the maintenance methylase would normally put a methyl group on a newly replicated bit of DNA. When the now hemimethylated DNA replicated, the strand that lacked the methyl group would give rise to another unmethyl-



TRANSMISSION of the DNA-methylation pattern in one cell to its daughter cells is accomplished by an enzyme known as a maintenance methylase (*circles*). The result of such transmission can be an identical pattern of protein binding and thus of gene activity and inactivity. When the DNA strands separate in preparation for replication, each one retains its original methyl groups, and the methylase quickly puts new methyl groups opposite them on the newly made strands (*orange*). The enzyme acts on doublets of the DNA bases cytosine (*C*) and guanine (*G*). The other bases are adenine (*A*) and thymine (*T*).



GENETIC SWITCHING that enables a precursor cell to give rise to one daughter cell with a changed gene-activity pattern and another daughter cell like itself can be accomplished in different ways. For instance, an enzyme might remove a methyl group from a specific base sequence of a newly duplicated gene, thereby causing the gene to become heritably unmethylated (*a*), or a specific binding protein might prevent methylation of the new strand (not shown). It is also possible that a specific enzyme may add a methyl group to a nonmethylated gene (*b*), which would soon become heritably methylated.

ated strand and would thereby erase the methylation signal from the affected chromosome.

These particular switching mechanisms may well be involved in the segregation of gene activities: the process in which a precursor cell divides to produce one daughter like itself and one that is committed to advancement along a differentiation pathway [*see illustration above*]. Segregation occurs in already determined stem cells in developing and mature organisms—for example, those in the bone marrow that give rise to a variety of red and white blood cells. Segregation events also occur quite early in development, and we proposed that these switches in gene activity might be based on methylation changes as well, although not necessarily on the same mechanisms that operate in stem cells.

Riggs's group and mine have also considered how the methylation of DNA might participate in the switching that early in development inactivates one of the two X chromosomes in every somatic cell of a female. (The process is random: in some cells it is the maternal X that is inactivated; in other cells it is the paternal X.)

There is a specific switch region known as the inactivation center in each X chromosome that has a role in determining whether the chromosome will be active. We proposed that this region may be recognized by a specific enzyme that methylates both strands of the inactivation center on the X chromosome destined to remain active but does not modify the homologous chromosome. This differential modification could occur if two conditions were met: the enzyme had low activity or had difficulty in methylating the target region, and as soon as the region was methylated, the event triggered a fast-acting reaction that blocked the other chromosome from being similarly altered. Thus, one chromosome would be methylated at the inactivation center and one would not. This pattern would be maintained faithfully through future cell divisions by means of the rules of inheritance set forth above.

A fuller explanation of the inactivation of an X chromosome also requires a mechanism for progressive inactivation of an entire chromosome early in development. This might be accomplished if the nonmethylated switch region of the chromosome that was destined to become inactive allowed a second methylase to attach to that site and to diffuse from there along the entire span of DNA, methylating specific base sequences.

Although I have discussed switching and heritability mechanisms as if they were controlled solely by factors within individual cells, it is well known that groups of cells in an embryo undergo coordinated changes in response to extracellular signals (known as inducers and morphogens) produced by other cells. The signals are received at the cell membrane and in some way are transmitted to the nucleus, where they influence the transcription of specific genes. Methylation could be a mediator of this group process, that is, the signal at the cell membrane might somehow alter the methylation state of the regulatory sequences in particular genes in a group of target cells. Once the methylation change was made, it would be inherited by subsequent generations of cells.

As I pointed out above, DNA methylation may not be the only way changes in DNA can be induced and made heritable. Yet as time goes by, the importance of our most basic assumption—that methylation is a component of gene regulation—is gaining experimental support. In many studies gene activation has been correlated with the absence of DNA methylation at specific sites, usually in or near the regulatory region of the gene. It is striking that this correlation is not a weak or relative effect but rather an all-or-none association: the presence of methyl groups seems to be associated not merely with a decline in gene expression but almost always with total inactivation. In contrast, alterations

in the base sequence of promoter regions, where transcription is initiated, often have a modulating effect, suggesting again that the methylation signal is the more specific regulator of gene function.

Such reports do not by themselves establish a cause-and-effect relation between methylation and the activities of genes, but other experiments are more persuasive. For example, when methylated and nonmethylated genes were prepared by recombinant-DNA methodology and introduced into cultured mammalian cells, only the nonmethylated genes were expressed. (These transfection experiments also provide evidence for the heritability of DNA-methylation patterns, since the foreign DNA retained its methylation state for many generations after it was taken up by the cultured cells.)

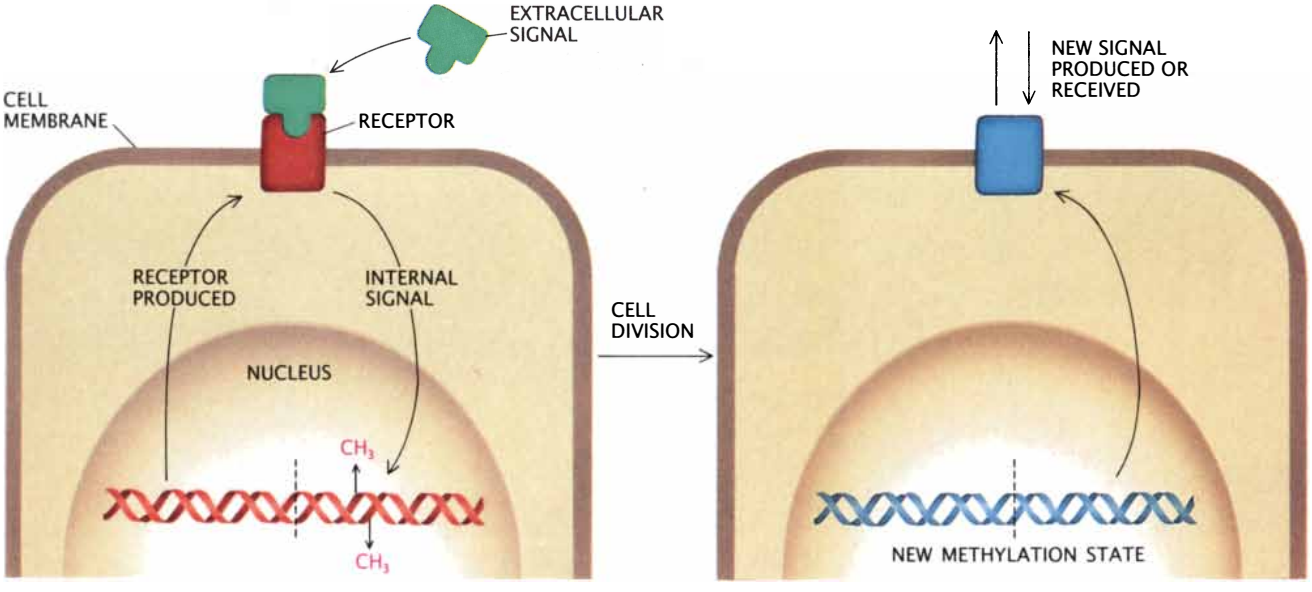
Various transfection studies have further shown that, when a methylated gene is introduced into a cell that normally expresses that gene, the inserted gene is thereupon demethylated and activated. This is true of the actin gene in myoblasts, cells that form muscle; actin is necessary for muscle contraction. It is also true for the insulin gene in cells derived from the pancreas and for an antibody gene in the immune system cells called lymphocytes. In contrast, when methylated versions of the genes are introduced into nonspecialized cells, they remain methylated and inactive. Related studies have demonstrated that the hormone-induced expression of the gene for vitellogenin (a component of egg yolk in amphibians and birds) is associated with the specific demethylation of the gene's regulatory region and that the gene for crystallin (an eye-lens protein) is demethylated in lens cells before being activated.

In all these cases the demethylation seems to take place in the absence of DNA replication, strongly indicating that the process is not a mere by-product of replication but is actively controlled by proteins that function only in specific cell types. The presence of such an active, tissue-specific process suggests that the demethylation is not an accident but exists for a purpose: selective gene activation.

More evidence that methylation influences gene activity has come from a totally different investigative approach: the chemical removal of methyl groups from DNA. Workers expose cells to a drug named azacytidine, which becomes incorporated into DNA and inhibits maintenance-methylase activity. As a result many sites that were originally methylated become unmethylated within a few cell-replication cycles. The drug has been shown in many experimental systems to reactivate the expression of genes that were previously silent, strongly suggesting that a loss of methylation is the cause. For example, inactive retroviral genomes that have been incorporated into cellular chromosomes are methylated, but they can lose their methyl groups and simultaneously become reactivated if they are exposed to azacytidine.

To cite another example, many mammalian cell lines fail to synthesize a particular active enzyme because they have a classical genetic mutation, that is, an alteration in the DNA sequence of the gene. Other cell lines that are enzyme-deficient actually have intact genes for the enzyme, but the genes are silent, probably because they have become methylated. When the cells are treated with azacytidine, from 10 to 30 percent of the formerly quiet genes become reactivated, an effect that represents about a millionfold increase over the spontaneous reversion rate. These experiments show that there can be very tight epigenetic control of gene activity and that the control is heritable.

Although the inactivation and reactivation of genes in cultured mammalian cells has provided valuable evidence for a relation between DNA methylation and gene expression, this does not by itself demonstrate that alterations of methylation patterns have a major role in the switching of gene activity during development. There is, however, suggestive evidence that active and inactive X chromosomes in female cells (which, it will be recalled, are made distinct during development) are methylated differently,



EXTERNAL SIGNALS can produce heritable changes in the activities of whole groups of cells during development, perhaps by influencing the methylation pattern of the DNA. The external signals are transmitted (*left*) by the binding of a diffusible protein (*green shape*) to a cell-surface receptor (*red shape*) (or by the binding of a protein on the surface of another type of cell). The signal-receptor complex might mediate a change in the methylation state of the cell's DNA, such as the removal of methyl groups from a gene, subsequently resulting in the production of a new receptor or a diffusible factor that can influence other cells. The change would be heritable and hence would be maintained in future cell generations (*right*).

THE BMW 5-SERIES.
THE PERFECT RECONCILIATION BETWEEN LOVE OF FAMILY AND PASSION FOR DRIVING.

The typical BMW engineer is a family man with a genuine passion for building and driving extraordinary automobiles.

So it's not surprising that BMW's 5-Series is the family-size four-door that "effectively recalibrates the performance sedan benchmark" (AutoWeek Magazine). This athletically-shaped BMW coddles the senses with rich leather seating, and ride comfort deemed the new standard in its class. And it reassures the mind with superior anti-lock disc brakes and a unitized steel body that is

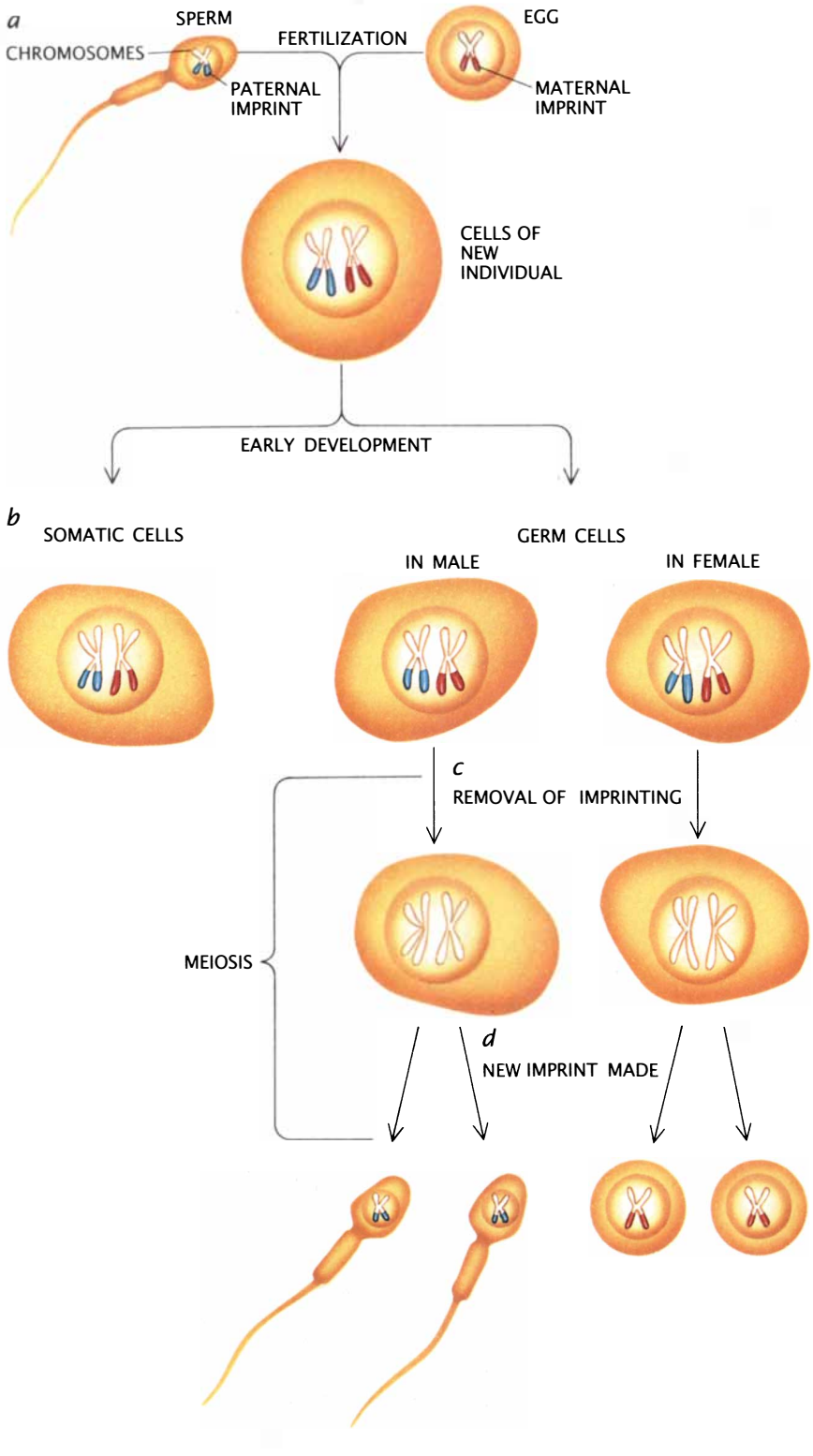
designed to absorb 35 percent more impact energy than Federal law requires.

But what this car does best is exalt the spirit. The sweet song of its in-line 6-cylinder engine, the feel of the road transmitted through its precise power steering, the pull of gravity as its 50/50 balance and patented independent suspension mate your tires to the curving pavement—all this must be experienced.

The experience is yours by contacting your BMW dealer and test driving the 5-Series sedan.

The car you need, reconciled with the car that you really want.

THE ULTIMATE DRIVING MACHINE.®



IMPRINTING of chromosomes, so that they are identifiable as having been donated to an individual (*a*) by the sperm of the father (*blue*) or by the egg of the mother (*red*), can persist for many cell generations in somatic cells (*b*). The imprinting, which seems to result in the differential activation and inactivation of maternal and paternal chromosomes, is altered, however, in germ cells, during meiosis. (The germ cells are set aside early in development and subsequently undergo meiosis to produce sperm or eggs.) The original imprinting is first erased (*c*) and then replaced by imprinting appropriate to the individual's sex (*d*). Thus, the chromosomes in the sperm are given a male imprint (even chromosomes passed to the male by his mother), and the chromosomes in the eggs are given a female imprint. Experiments suggest imprinting is accomplished by the reversible methylation of chromosomes.

just as our model and Riggs's postulated might be the case. In particular, certain DNA sequences in or near the regulatory regions of "household" genes have been found to be methylated in the inactive X chromosome but not methylated in the active homologue. Household genes, which are responsible for cell maintenance, are normally turned on in all cells. Genes on the inactive X chromosome can also be reactivated by azacytidine.

Even more direct is the work of Peter A. Jones and Shirley M. Taylor of the University of Southern California School of Medicine at Los Angeles. They discovered that incompletely differentiated fibroblasts, which can be regarded as connective-tissue cells, could be converted into other specialized cell types by treatment with azacytidine. Some cells became myoblasts (muscle-cell precursors) and then mature muscle cells that actually twitched in the cell-culture plate. Other cells in the population of fibroblasts differentiated into fat-storing cells known as adipocytes. It has been shown that the differentiation of the muscle cells was associated with the loss of methylation in a particular regulatory gene, which would indicate that the methylation state of genes can indeed influence the developmental pathways of cells.

If somatic cells become determined, differentiated and able to give rise only to certain cell types during development, what enables germ cells to remain totipotent, so that when an egg is fertilized by a sperm cell the resulting embryo can give rise to every cell type in the new individual?

Apparently germ cells do not follow the unidirectional course of somatic cells. Any epigenetic changes in the germ cells, such as the inactivation of an X chromosome, are reversed or altered during or just before meiosis, the process of cell division that gives rise to sperm and eggs. Also, the chromosomes in the sperm and egg are differentially imprinted, or marked as having come from the father or the mother, respectively. Hence, even if chromosomes in, say, a sperm happen to have been inherited from the father's mother, those chromosomes would still bear the male imprint. Such imprinting has been found to be crucial to normal development in mice (and, by inference, in other mammals), which implies that the imprinting silences and activates different sets of genes in the maternal and paternal chromosomes and that the differenc-

The idea of technical excellence at Alcoa is part and parcel of a larger imperative: superior quality and value through technology. In this spirit, our 1989 Technology Awards go to 37 Alcoans who have carried technological discoveries through to their full realization—setting new standards for excellence.

37 Alcoans Who Weren't Content To Let Well Enough Alone

The 1989 Francis C. Frary Award

for a career of outstanding individual contribution to Alcoa technology:

Dr. Joseph T. Laemmle

Manager, Surface Technology Division, Alcoa Laboratories

In a career of applying science to solving plant-floor manufacturing problems, Joe Laemmle has become Alcoa's authority on tribology—the science of friction, lubrication, and wear. His work has pointed the way to major improvements in lubricant systems for rolling, forging, and extruding aluminum.

The 1989 Arthur Vining Davis Award

for outstanding group achievements in Alcoa technology:

Two teams of technologists have earned this award for 1989. At Alcoa Lebanon (Pa.) Works, 21 Alcoans developed a method for changing the width of roll-cast aluminum sheet without stopping the casting process. Their patented innovation has improved uptime, recovery, and capacity in the continuous casting of sheet.

Team members:

Larry L. Bishop
Kenneth H. Daullary
Thomas W. Ebright
Donald D. Finkel
Michael R. Garrison
William E. Harris
Dennis P. Heagy
William H. Hoffman
James Kosh, Jr.
James T. Longenecker
Larry G. McKinney
Leonard T. Miller
Hilary J. Picard
James G. Riegle, III
David R. Smith
Earl E. Smith
Edwin W. Smith
Harold K. Stoner
Frank P. Swigon
Janie L. Waybright
Richard E. Wenrich

In a cooperative effort spanning languages, cultures, and scientific disciplines, a group of five U.S. Alcoans and six Brazilian Alcoa Aluminio S.A. employees worked out a commercial process to make high alpha alumina chemicals in a fluid flash calciner. The new technology lowers processing costs and can be installed on any flash calciner at almost no capital cost.

Team members:

Nestor Andrade
Ricardo de Carvalho
Ronald G. Chabal
Terrance R. Clever
João Carlos da Silva
Roberio Farias
Paulo Ferreira
Amarildo Goncalves
John W. Kastner
Alan Pearson
Richard B. Tallarico

The 1989 Chairman's Award

for significant contributions to the development and implementation of materials, process, and systems technologies:

Thomas F. Drumwright, Jr.

Known as an international expert in non-destructive evaluation of materials, Tom Drumwright has been a force for meeting and increasing product quality standards throughout Alcoa. In the process, he has helped to set inspection standards for the entire industry.

James E. Knepp

Jim Knepp's expertise in lubrication technologies has given Alcoa a significant competitive advantage in rolling aluminum. Known for moving laboratory discoveries into production—for Alcoa and its customers—he holds numerous U.S. and foreign patents.

Dr. Jan H. L. Van Linden

A 1986 Davis Award recipient for his role in recycling technology, Jan Van Linden has spearheaded the transfer of this technology from the laboratories to Alcoa plants where beverage containers are recycled. His work in reducing metal loss in the recycling process has achieved sizable cost savings.

Robert W. Westerlund

At our Davenport (Iowa) Works, Bob Westerlund has helped to modify heat treating operations, contributing to Alcoa's lead in high-strength aluminum aircraft alloys. A 1987 Davis Award winner, Bob has also coordinated the development of lightweight aluminum lithium alloys.

ALCOA

es complement each other in directing normal development.

An "epigenetic code" must, then, be superimposed on the normal DNA of eggs and sperm if development is to be initiated and carried through properly. Recent work indicates—perhaps not surprisingly—that this reprogramming of the genome at meiosis may well involve changes in methylation patterns: in mice the chromosomes derived from sperm and eggs are differently methylated.

Regardless of what controls embryonic development in higher organisms, the result is a fully formed organism whose body generally remains quite stable for a fairly long time. Yet ultimately the body ages; tissues gradually function less efficiently. One probable contributing cause is the gradual failure of cell-maintenance mechanisms such as the ones that repair DNA, replace lost cells and regulate gene transcription.

If the heritability of methylation patterns by cells is important to the long-term functioning of tissues, then a gradual loss of methyl groups could contribute to aging. For example, there may be occasional failures of maintenance methylation during mitosis or a failure to replace methyl groups after repair of DNA in nondividing cells. The loss of methylation might then lead to the undesirable activation of silent genes and thus to derangements of cellular function and to cell death.

This scenario would help to explain two recent findings in mice. A gene on the inactive X chromosome is reactivated at a low frequency in young animals but at an increasingly higher frequency as the animals age. Similarly a gene necessary for pigment formation is silenced when it is inserted into an inactive X chromosome; the result is an albino coat. The gene is progressively activated as the recipients age, and the coat displays increasing pigmentation.

Experimental evidence from human cells supports the loss-of-methylation scenario. Leonard Hayflick, working at the Wistar Institute of Anatomy and Biology in Philadelphia, discovered that the life span of human cells in culture depends on the number of times cells divide, not on chronological time. In such cells it has been shown that the number of methylated sites on DNA declines with increasing numbers of cell divisions and that the cells eventually lose their ability to proliferate. The likelihood that this age-related inability to multiply is at least in part a consequence of the loss of methylation is suggested by experiments in which cultured cells are treated with azacytidine. A single treatment of young cells, which significantly reduces the overall level of DNA methylation, initially has little effect on the functioning or growth rate of the cells. Yet the cells apparently "remember" the exposure; they die out well before untreated controls.

The work on aging is not yet conclusive, but it does indicate that epigenetic controls on gene expression are not absolute and can deteriorate. Similarly it is conceivable that epimutations (aberrations in gene activity) might cause cells to become "dedifferentiated" and might thus contribute to carcinogenesis.

Among the reasons for suggesting that epimutations might be important in cancer is the finding that metastatic cells arise from nonmetastatic ones more frequently than would be expected if genetic mutation alone were the cause. Also, mutation rates seem to be similar in mice and human beings, and yet human cells are much less likely to be transformed into cancerous ones either spontaneously or after exposure to carcinogens. This finding indicates that the difference in transformation rates may be due to stronger epigenetic controls in human cells.

So far, changes in the methylation of DNA have not been directly implicated in carcinogenesis, but there is some indirect evidence. For example, azacytidine is a highly potent carcinogen, producing a wider range of tumor types in experimental animals than is produced by other well-known carcinogens. Yet azacytidine does not generally cause genetic mutations. Given that the chemical is known to affect DNA methylation, it seems reasonable to suspect that the azacytidine-caused alterations in methylation patterns contribute to cancer.

Collectively, then, the data point to methylation of DNA as an important epigenetic factor in the development of higher organisms. The data also suggest a possible role for the failure of normal methylation processes in cellular aging and in the transformation of normal cells into cancerous ones.

The potential importance of cytosine methylation in cell processes is underscored by the fact that it is found in a wide range of organisms, and yet it may not be universal. Indeed, the nematode (roundworm) *Caenorhabditis* and the fruit fly *Drosophila,* which undergo complex processes of development, are reported to have no 5-methylcytosine in their DNA.

The Tandy® 1000 SL

Now only $699

Save $200 on the easiest-to-use PC compatible made in America—just power and run!

We've made it easy and affordable to get into PC-compatible computing. Now you can get the Tandy 1000 SL at our lowest price ever. Buy a complete system with CM-5 color monitor for just $998.95.

Most PC compatibles charge extra for the operating system. With the 1000 SL, MS-DOS® is not only included, it's actually *built into memory.* Seconds after you turn the 1000 SL on, MS-DOS has loaded, and up and up comes the DeskMate® Graphical User Interface—a neat, organized display of your programs. This simple, plain-English approach to computing is the ultimate in ease of use—and a Tandy exclusive.

The 1000 SL even comes with DeskMate productivity software. With DeskMate, you're ready to go with ten popular applications that let you write letters, prepare budgets, file information, draw colorful pictures and more. There's even a 90,000-word spell checker, address book, Hangman word game and the PC-Link® online information service.

And listen to this: Tandy offers the only PC compatibles that let you make audio recordings onto your own 5¼" diskettes.

Come into Radio Shack today and get the Tandy 1000 SL at this terrific low price. See just how easy personal computing can be.

Tandy Computers: Because there is no better value.™

Radio Shack®
The Technology Store™
A DIVISION OF TANDY CORPORATION

Sale begins 3/19/89, ends 6/19/89. Prices apply at Radio Shack Computer Centers and participating stores and dealers. Tandy 1000 SL reg. $899. Monitor sold separately. Modem required for communications; microphone required for recording. MS-DOS/Reg. TM Microsoft Corp. PC-Link/Reg. TM Quantum Computer Corp.

"Maybe the problem really wasn't any of my business. But I'd noticed we were spending too much time hand-forming the mounting frames for our rockets.

I suggested it would be a good idea to spin the frames, because I could do them a lot faster that way. Spinning is an art. A way of working metal that gives you more control. It's what I specialize in. Well, the management here realizes the guys on the line know what they're doing; so they gave me the go-ahead.

Now we've cut the frame-making time down from an hour and a half to just seven minutes. And the quality is perfect. Never any rejections. You can't beat that." *—Austin Varney, Delta Rockets, Metal Spinner*

MCDONNELL DOUGLAS
A company of leaders.

This finding has led some investigators to conclude that cytosine methylation is not important in unfolding the program of development in higher eukaryotes.

Their point is well taken, but there are other possible explanations. Perhaps *Caenorhabditis* and *Drosophila* do in fact depend on 5-methylcytosine for genetic regulation but the amount present in their DNA is undetectable by measuring techniques currently being used. Only time and more research will tell whether the epigenetic controls over development in different species are based on the same molecular mechanisms or on several different ones.

Whichever factors turn out to be important in controlling the development of *Caenorhabditis* and *Drosophila,* there is no denying that methylation of DNA seems to be an important epigenetic factor in many eukaryotes. Yet methylation cannot explain every feature of development. The controls over cell behavior and characteristics exert their effect at several levels: interactions of DNA and protein, signal transduction from cell membrane to chromosomes and communication between cells. A broad unifying theory of development is needed if the field of epigenetics is ever to explain the behavior of genes during that process with as much clarity as Mendelian genetics has explained the transmission of genes from generation to generation. Such a framework is needed for the formulation of testable predictions and for the continual interplay of theory and experiment that is essential to the advancement of science.

FURTHER READING

5-METHYLCYTOSINE, GENE REGULATION, AND CANCER. Arthur D. Riggs and Peter A. Jones in *Advances in Cancer Research,* Vol. 40, pages 1–30; 1983.

THE MOLECULAR BIOLOGY OF DNA METHYLATION. R. L. P. Adams and R. H. Burdon. Springer-Verlag, 1985.

AGEING: X-CHROMOSOME REACTIVATION. Robin Holliday in *Nature,* Vol. 327, No. 6124, pages 661–662; June 25, 1987.

GENOMIC IMPRINTING: MEMORIES OF MOTHER AND FATHER. Marilyn Monk in *Nature,* Vol. 328, No. 6127, pages 203–204; July 16, 1987.

THE INHERITANCE OF EPIGENETIC DEFECTS. Robin Holliday in *Science,* Vol. 238, pages 163–170; October 9, 1987.

TRANSGENES AS MOLECULAR PROBES FOR GENOMIC IMPRINTING. M. Azim Surani, Wolf Reik and Nicholas D. Allen in *Trends in Genetics,* Vol. 4, No. 3, pages 59–62; March, 1988.

DNA METHYLATION AND GENE ACTIVITY. Howard Cedar in *Cell,* Vol. 53, No. 1, pages 3–4; April 8, 1988.

SCIENTIFIC AMERICAN CORRESPONDENCE

Offprints of more than 1,000 selected articles from earlier issues of this magazine, listed in an annual catalogue, are available at $1.25 each. Correspondence, orders and requests for the catalogue should be addressed to W. H. Freeman and Company, 4419 West 1980 South, Salt Lake City, Utah 84104. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Manuscripts are submitted at the authors' risk and will not be returned unless accompanied by postage.

Advertising correspondence should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Address subscription correspondence to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 3187, Harlan, IA. 51593. Telephone inquiries: 1-800-333-1199, U.S. only; other 515-247-7631/32. The date of the last issue on subscriptions appears on each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name

New Address	**Old Address**
Street	Street
City	City
State and ZIP	State and ZIP

The Fossils of Monte San Giorgio

An ancient sea provides a rich assemblage of vertebrates from the Triassic period

by Toni Bürgin, Olivier Rieppel, P. Martin Sander and Karl Tschanz

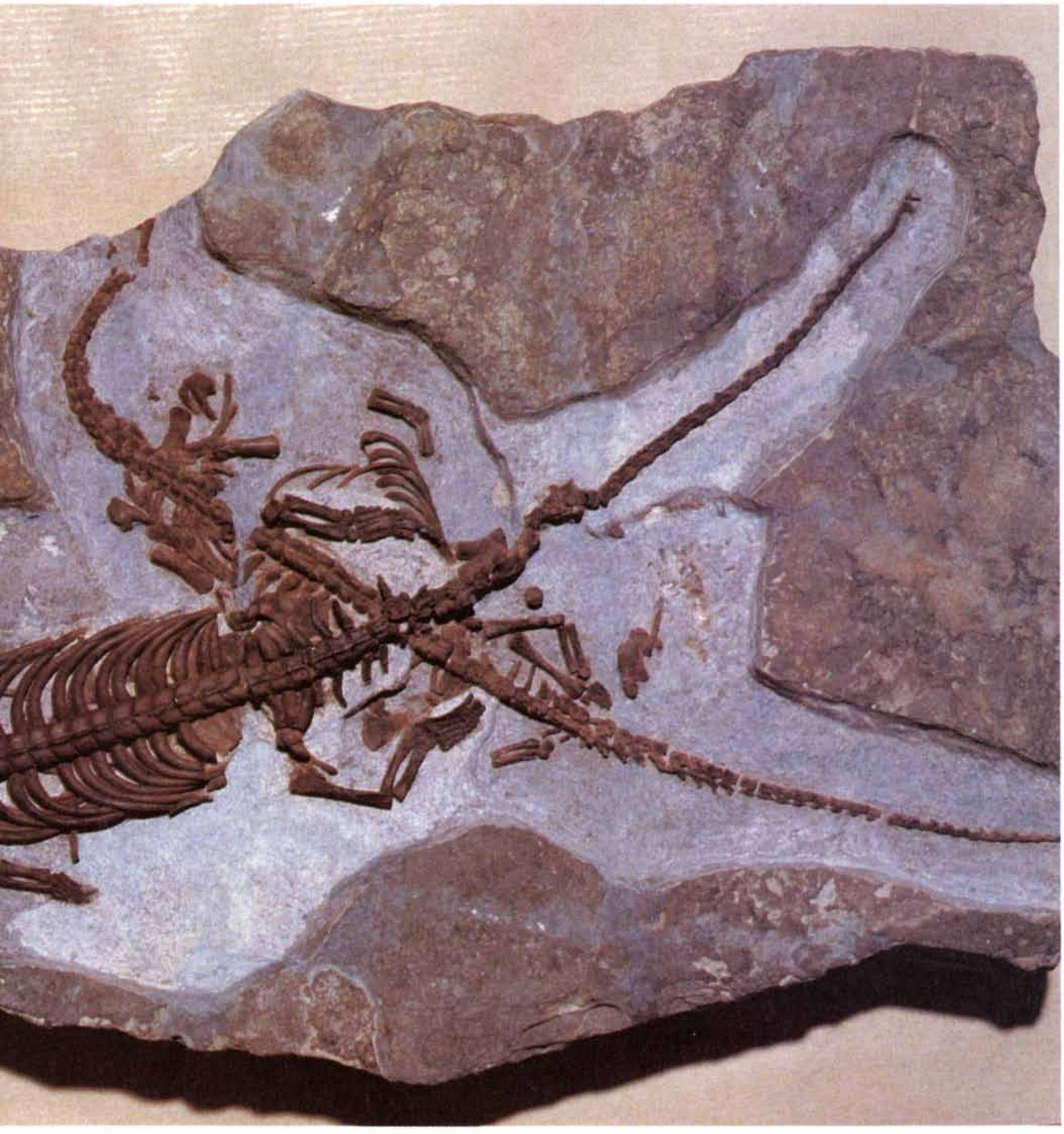
A little more than 200 million years ago, in a part of the world that is now the border between Switzerland and Italy, there existed a small sea. Although this body of water eventually dried up and disappeared, the remains of its animal inhabitants—a diverse assemblage of fishes and reptiles—were preserved after their death in the muddy sediments at the bottom of the basin. Over the course of millions of years the sediments hardened, were uplifted by tectonic activity and eventually were transformed into a mountain called Monte San Giorgio. Today Monte San Giorgio, which rises above Lake Lugano in the southern Alps, is something of a tourist attraction, not so much for its geological past as for the panoramic view it offers those who succeed in reaching its peak.

Yet it is the richness of the fossils embedded in its sedimentary rock that gives the site its real distinction. The thousands of fossils that have been excavated from Monte San Giorgio over the years have provided paleontologists with an unparalleled, if somewhat puzzling, view of the past. The fauna represents a pivotal time in the history of life on earth known as the Triassic period, which lasted from 210 to 250 million years ago. It was during this relatively short time span that several major groups of terrestrial vertebrates made their debut, including the ancestors of most of the modern reptiles, the dinosaurs, the mammals and possibly the birds.

The first fossils from Monte San Giorgio were discovered in 1863, when Antonio Stoppani, then director of the Natural History Museum of Milan, stumbled on several vertebrate fossils (some fishes and a few reptiles) at the base of the mountain near Besano, a small village in northern Italy. On returning to Milan, he gave the fossils to Francesco Bassani, who published brief descriptions of them and classified them taxonomically. Although Stoppani returned to the site again in 1878 and collected several more specimens, the brief anecdotal accounts of his findings stimulated scant interest. It was not until the turn of the century, some 40 years later, when the bituminous shales (sedimentary rocks rich in organic matter) of Monte San Giorgio were exploited for an oily compound called saurol (believed to be an effective remedy for rheumatism), that the area's fossils were once again the focus of attention.

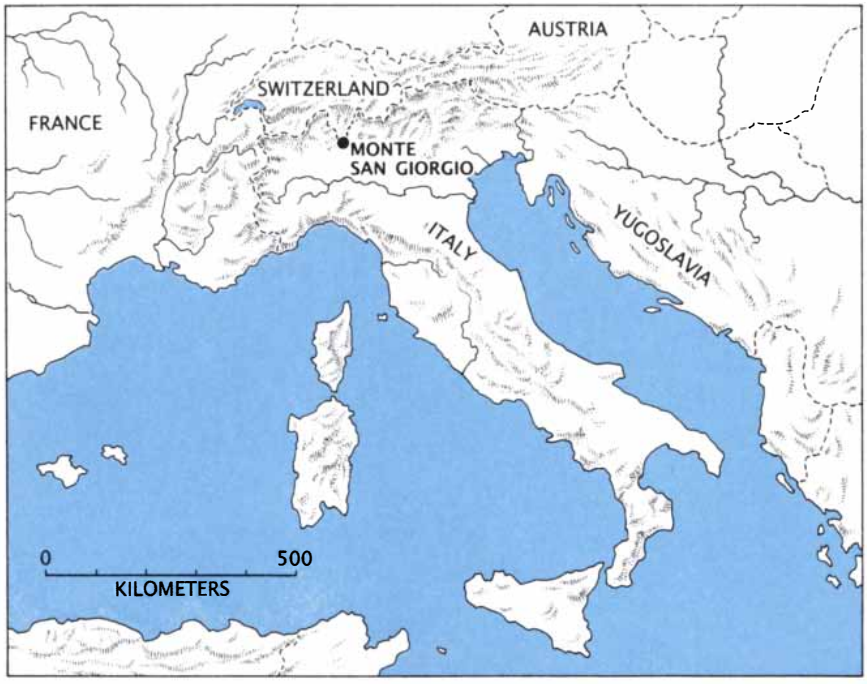
The year was 1919 and Bernhard Peyer, a young zoologist from the University of Zurich, was skimming through a pile of discarded shale near the saurol extraction site; as he sorted through the pieces, his hammer struck a piece of shale on which was preserved the paddlelike limb of an extinct marine reptile called an ichthyosaur. No comparable fossil of this age had been found elsewhere in Europe, and not surprisingly, Peyer's discovery generated a great deal of excitement. Spurred by the importance of his find, Peyer devoted much of the rest of his life to a systematic search for addi-

TONI BÜRGIN, OLIVIER RIEPPEL, P. MARTIN SANDER and KARL TSCHANZ, listed here in alphabetical order, are members of a fossil vertebrate study group at the Paleontological Institute and Museum at the University of Zurich. Bürgin, who has a Ph.D. from the University of Basel, is currently studying the fishes of Monte San Giorgio for the Swiss National Science Foundation. Rieppel, whose Ph.D. is also from the University of Basel, is a privatdozent, who teaches vertebrate paleontology and supervises research on the vertebrates of Monte San Giorgio. Sander has an M.A. from the University of Texas and a Ph.D. from the University of Zurich. Tschanz has a Ph.D. from the University of Zurich.



EXTINCT REPTILE, the nothosaur* Serpianosaurus, *is present in large numbers in the shales of Monte San Giorgio. The animal superficially resembled a lizard but was amphibious, spending some of its time in the water and some on land. This species, which reached about 70 centimeters in length, had a long, flexible neck and probably fed on small fish. Near the base of its tail, and at right angles to it, is the backbone of another individual.

tional fossil material from the region. When he retired in 1956, the task of collecting and cataloguing fossils was passed to his successor Emil Kuhn-Schnyder, who in turn added thousands of new specimens to the collection at the University of Zurich. Today, largely because of the efforts of these two men, the Paleontological Institute and Museum at the University of Zurich is considered to have the world's finest collection of Triassic marine



MONTE SAN GIORGIO rises near Lake Lugano in the southern Alps, not far from the border between Switzerland and Italy. Most of the fossils come from exposed beds on the southern slope of the mountain, north and northwest of the Swiss villages of Meride and Serpiano.

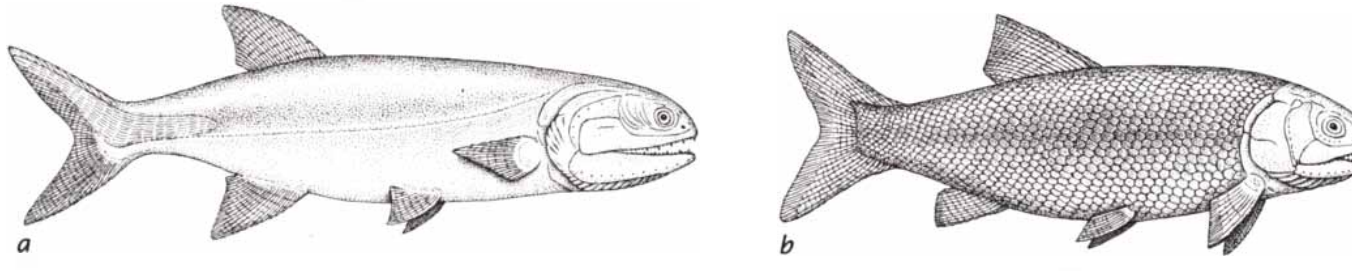
vertebrates and is home to more than 4,500 specimens from Monte San Giorgio alone.

By analyzing both the nature and stratigraphy of its sediments, geologists have determined that in the Triassic period the San Giorgio basin was not very big: it is estimated to have been from six to 10 kilometers in diameter and approximately 100 meters deep. Careful analysis of fossil and stratigraphic data indicates that the basin was not far from the coast of an ancient ocean called the Tethys Sea, a predecessor to the now existing Mediterranean. Although the basin itself was surrounded by algal reefs and dry land, the presence of pelagic (open-sea) species indicates that it was accessible, at least intermittently, from the open sea.

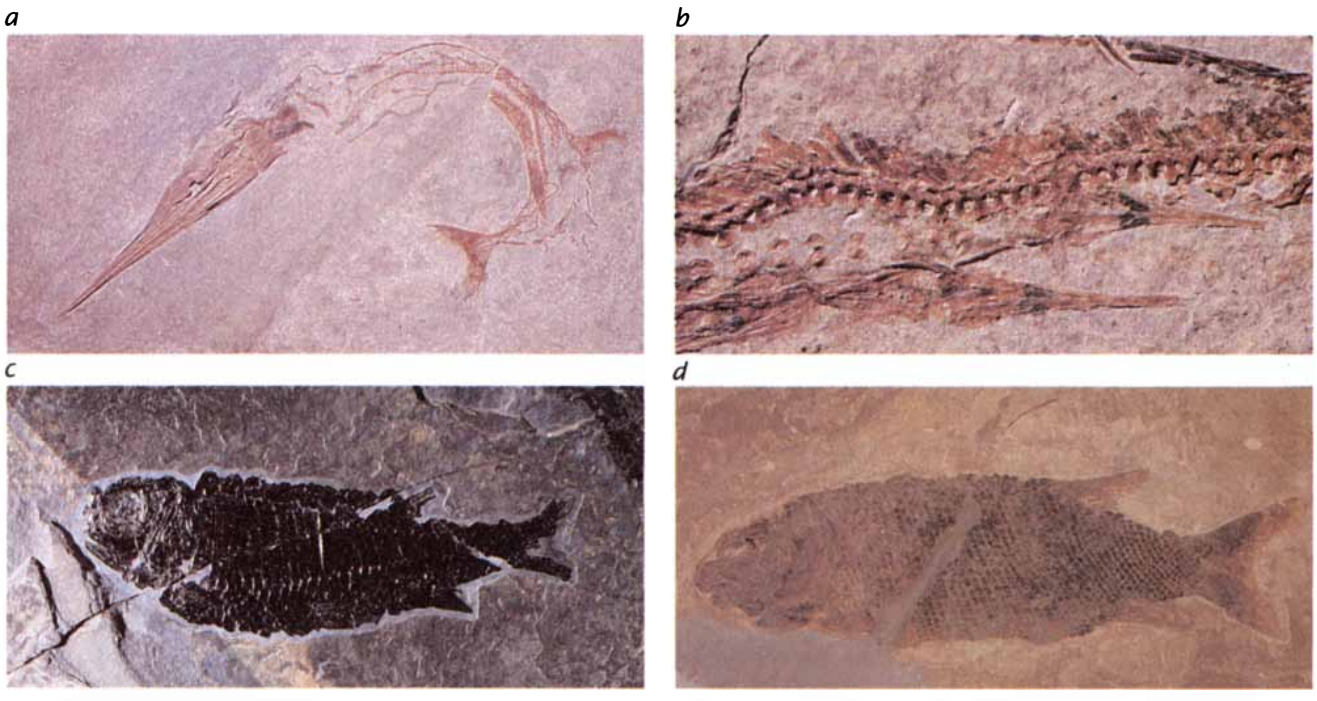
According to Hans Rieber, currently the director of the Paleontological Institute, the bottom of the basin consisted of fine-grained mud; when the animals that lived in the basin died, they sank to the bottom, where conditions were anoxic (without oxygen), and so their remains, which would normally be broken down by aerobic bacteria and other scavengers, were protected from decay. Over the course of many thousands of years the mud accumulated and turned to stone, compressing and flattening the animals' skeletons as it petrified. In some instances the force of compression crushed the skeletons so severely that interpretation of fine anatomical detail is difficult, if not impossible. Most of the fossils are well preserved, however, and even 200 million years after their death delicate bones and fine details such as tiny spines and scales are distinctly visible.

It is readily apparent—even at the glance of an untrained eye—that the San Giorgio basin supported a wide variety of animal species. In terms of fish diversity alone, the basin can be considered analogous to a modern coral reef. Five species of sharks have been identified so far; four of them are small, yet robust in shape, with crushing teeth that suggest a diet of shellfish. The fifth species, represented by a rather large fin spine and some teeth, is estimated to have been from two to three meters long and to have fed on shellfish. Like fossil sharks from other localities, the San Giorgio specimens are represented mostly by teeth and an occasional fin spine or backbone (sharks, unlike bony fish, have skeletons made of cartilage and deteriorate rapidly after death). Yet several almost complete shark specimens have been found, providing further testimony to the exceptional preservation of the site.

Lobe-finned fishes are present in limited numbers, but the specimens that have been found tend to be complete because their heavy, enamel-like scales resist decay. These fishes, members of a mostly extinct order called the Actinistia (the group to which the living coelocanth *Latimeria* belongs), have been the focus of much attention because they are considered to be related to the stem group from which all land vertebrates descended. So far more than 20 specimens have been collected from the San Giorgio shales, all of which fall into one of three distinct size categories. Such



DIVERSE ARRAY OF RAY-FINNED FISHES has been excavated from the bituminous shales of Monte San Giorgio. More than 550 specimens in some 30 genera have been collected so far. Representatives, shown here drawn to scale, fall into three distinct size classes. In one category are the large predatory fishes, such as **Birgeria stensiöi (a)**, *which grew to more than a meter in length,* **Colobodus bassanii (b)**, *which was about 70 centimeters (or 27.5 inches) long, and* **Saurichthys cu-**

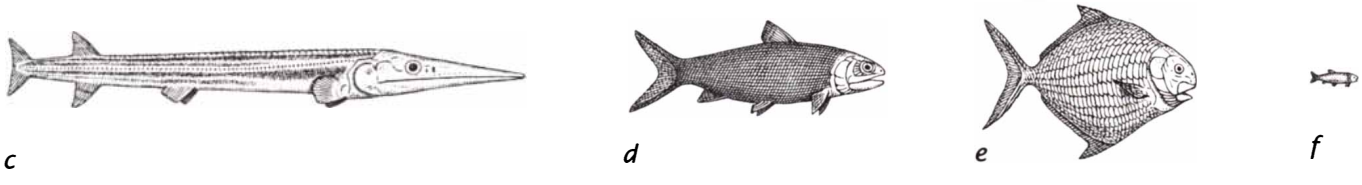


EXQUISITE PRESERVATION of the San Giorgio fossils is exemplified by some of the ray-finned fishes. Fine details such as the tail fins and elongate snout of the lizard fish Saurichthys *are visible* (a)*; even more remarkable is the presence of two embryos inside the abdomen of a female* (b)*, evidence that these fish gave birth to live young. Just below the tail of the small, heavily armored* Peltopleurus (c)*, a highly modified anal fin can be seen. The fin bears vertical hooklike rays that may have played a role in courtship by enabling males to hold onto females. One, as yet undescribed, species* (d) *is informally called "big teeth" in reference to the animal's unusual dentition; its diet remains a mystery.*

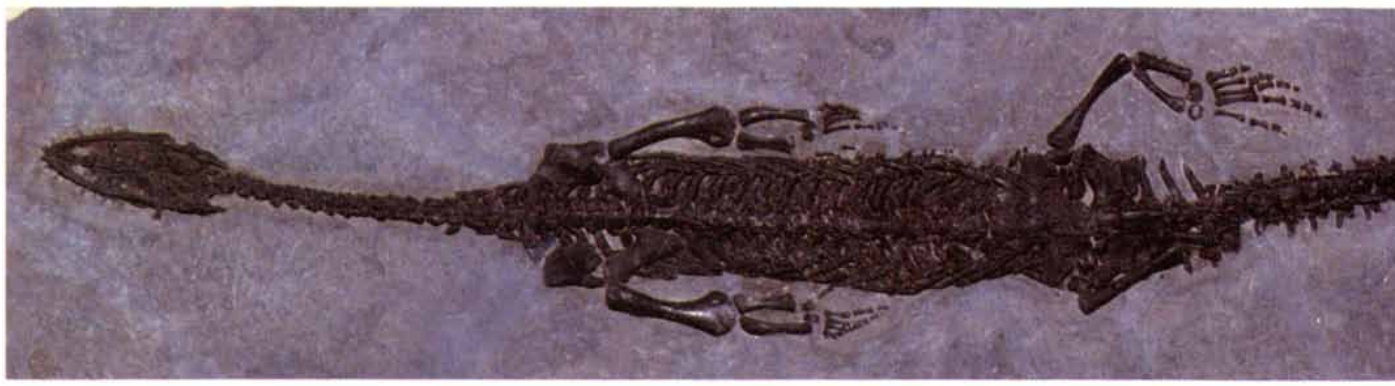
discrete size distribution leads us to speculate that the size groupings may reflect ecological—perhaps dietary—specialization. The largest ones measure about 70 centimeters in length and may have fed on small fish. In contrast, the smallest ones were only 20 centimeters long, but their eating preferences are not known.

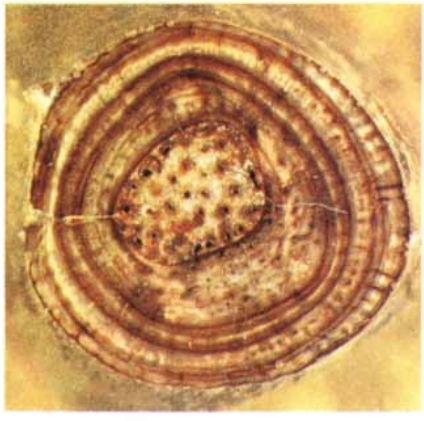
More than 550 well-preserved specimens of ray-finned fishes (the group to which the majority of living fishes belong) have been catalogued. Although most have yet to be studied in great detail, we expect that once they are carefully studied, they will shed new light on the overall evolution and relationships of ray-finned fishes. Already these fishes have provided valuable information about the paleoecology of the San Giorgio basin. The presence of ocean-dwelling, fast-swimming forms, for example, provides some of the evidence that the basin was connected intermittently to the Tethys Sea. One such indicator fish is *Birgeria,* which was widespread during the Triassic and is common in many deposits of that period.

In contrast to the fishes, which have become the focus of intense study only recently, the reptiles of San Giorgio have been the objects of paleontological curiosity since Peyer published his first account of them. The most abundant and perhaps best studied are the amphibious nothosaurs, members of a large and common order of Triassic reptiles called the Sauropterygia. Nothosaurs are descendants of a branch of terrestrial sauropterygians and retain many of the features of their land-dwelling relatives; their limbs, for example, were not specially adapted for aquatic locomotion. Unlike many aquatic reptiles, including ichthyosaurs, whose limbs were modified into paddles and therefore played an important role in propelling the animals through water, nothosaurs are thought to have moved forward by lateral undulations of the trunk and tail. In addition to having elongate, flattened tails, most had long, flexible



rionii (c)*, approximately 60 centimeters long. In another category are the medium-size fishes such as* Ptycholepis barboi (d)*, which was about 30 centimeters long and superficially looked like a modern mackerel, and* Bobasatrania (e)*, an inhabitant of coral reefs. In the smallest category are fishes such as* Peltopleurus (f) *that were five centimeters or less in size; they are preserved in the shales of Monte San Giorgio in large numbers.*



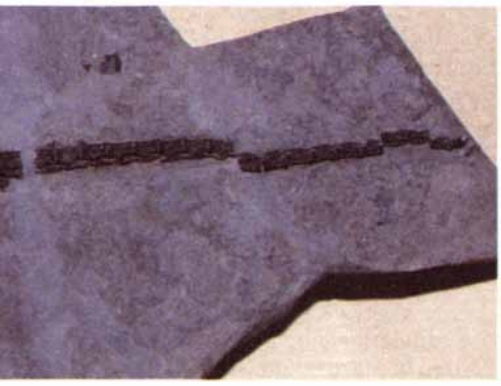


SMALLEST yet most abundant reptile at Monte San Giorgio is **Neusticosaurus (above).** ***The animal's feet are better adapted for paddling through water than for walking on land. A thin cross section of a*** **Neusticosaurus** ***leg bone*** **(left)** ***reveals the presence of annual growth rings. The animal from which this section of bone was taken was five years old when it died; the diameter of the bone is 4.1 millimeters.***

necks and were probably adept at catching small fish as they swam.

The San Giorgio nothosaurs varied in shape and size—from *Ceresiosaurus,* an animal that grew to as much as three meters in length, to *Neusticosaurus,* the dwarf lizard, which averaged only about 30 centimeters (a foot) in length. Altogether some 400 specimens—representing each stage in the life cycle, from embryo to adult—have been documented. Development in these animals can be studied because nothosaurs, like all cold-blooded reptiles, grew by adding new bone in the form of annual growth rings. Thin cross sections of their bones (about 50 micrometers thick) can be examined under a microscope and the number of rings counted. Extensive analysis of the growth rings (made possible by the sheer number of nothosaur specimens in the Institute's collection) reveals that the animals became sexually mature at about three or four years of age and lived to a maximum of six years.

Related to the nothosaurs but fully aquatic are the placodonts: a group of short, stout marine reptiles that had large, flattened teeth and are believed to have fed primarily on bivalves and other mollusks found along the edge of the basin. The two genera of placodonts known from Monte San Gior-



gio are *Cyamodus* and *Paraplacodus*.

Another well-represented group of reptiles at San Giorgio is the Archosauromorpha. The fossils in that category constitute a large and poorly understood assemblage but are of interest because they are members of the group that eventually gave rise to the dinosaurs. One of the more bizarre archosauromorphs—indeed perhaps the strangest reptile ever described—is *Tanystropheus*. Known as the giraffe-neck saurian, *Tanystropheus* has become something of a mascot for the San Giorgio fossil fauna. The animal, which measures as much as 4.5 meters from head to tail, is famous for its absurdly long neck, which is more than twice the length of its trunk. We have concluded that the animal must have been aquatic because it is impossible to think that it could have supported the weight of its neck on land.

Closely related to *Tanystropheus* but much smaller in size (no more than 80 centimeters from head to

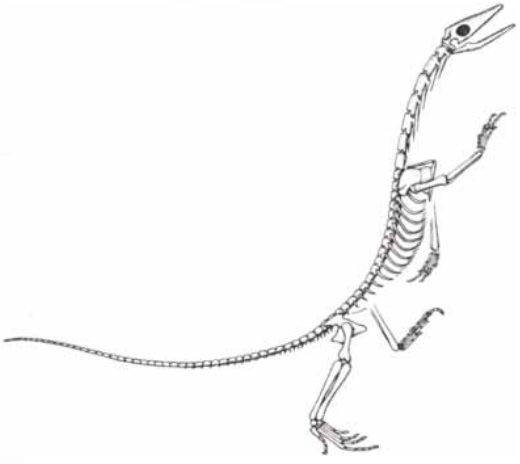
NOTHOSAUR* Ceresiosaurus, *shown here surrounded by eight small* Neusticosaurus *specimens, is exquisitely preserved. It is 2.3 meters (7.5 feet) long and was the best adapted among the nothosaurs to an aquatic habitat. Its long feet were somewhat paddlelike, and its tail was large and laterally flattened for better propulsion through the water.

tail) was *Macrocnemus*. The animal is thought to have been fairly mobile on land, where it may have spent considerable time chasing insects and grabbing them with its mouth. The large number of specimens of different age classes of both *Macrocnemus* and *Tanystropheus* indicates that their grossly elongated necks are a function of allometry, that is, the necks grew at a rate faster than the rest of the body. It is difficult to believe that the exaggerated neck of *Tanystropheus* could have provided its bearer with any kind of advantage; indeed, we think that the animal survived for several million years not because of its extraordinary neck but in spite of it.

The only true archosaur (the group to which dinosaurs belong) found at the site is *Ticinosuchus ferox*. The animal, which was about 2.5 meters long, is believed to have been terrestrial, and a ferocious carnivore.

Thalattosaurs, another enigmatic group of marine reptiles, are also represented in the San Giorgio shales. These animals belong to the Diapsida, the group from which all living reptiles, except turtles, descended. The thalattosaurs are a poorly known group of marine reptiles. Three genera from San Giorgio have been described: *Askeptosaurus,* a fish-eating animal that measures about 2.5 meters in length and has a long, narrow skull; *Clarazia,* about one meter long with crushing teeth (with which it probably fed on mollusks); and *Hescheleria,* an incomplete specimen whose dietary habits are completely unknown.

Ichthyosaurs, a group of marine reptiles that were similar in size and shape to modern dolphins, having paddle-shaped limbs and distinctive snouts, were also inhabitants of the San Giorgio basin. Abundant and widely distributed during the Triassic, they are nonetheless a puzzling group; their relationship to other reptiles, for example, is still highly conjectural. Some specimens have been found that con-

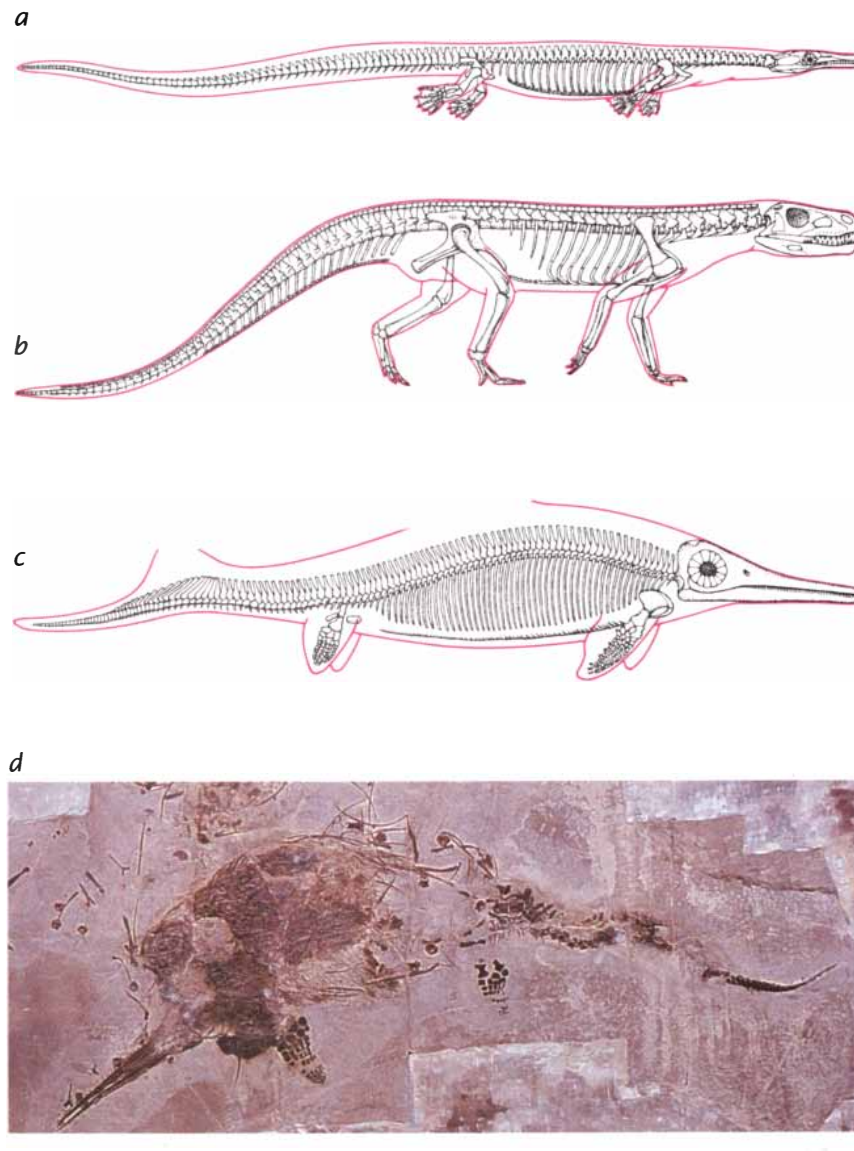


***BIZARRE REPTILE** from the San Giorgio basin was* **Tanystropheus**, *a 4.5-meter-long animal, whose neck was twice as long as its trunk. The relative weight of the neck was so great that the animal could probably have supported it only in water. For that reason the species is thought to have been aquatic. The head, which is missing from the fossil* **(above)**, *is known from other specimens and was unusually small relative to the rest of the body* **(below)**. *Related to* **Tanystropheus** *but much smaller was* **Macrocnemus (above right)**, *a highly mobile reptile that probably spent considerable time on land and may even have assumed an upright position while hunting insect prey.*

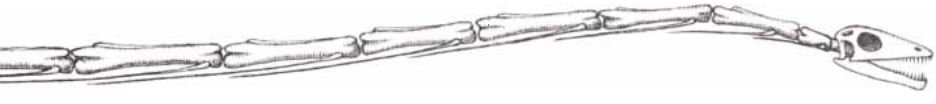
tain the remains of unborn young; it is surmised that the eggs must have developed within the mother's body and that the young were born live, a strategy that seems clearly advantageous for an aquatic reptile unable to come out on land to lay its eggs. Three genera of ichthyosaurs have been found at San Giorgio: *Mixosaurus,* the most common and best known of Triassic ichthyosaurs, *Cymbospondylus,* and possibly *Toretocnemus.*

In summarizing the fossils of Monte San Giorgio it can be said that they offer an exquisite glimpse at a now extinct fauna. Yet the anatomical similarities and confusing traits of many of the reptiles make it difficult to fit them into the framework of existing classifications. The diversity of the San Giorgio vertebrate fossils seems only to enforce the prevailing notion that relationships among the major groups of fishes and reptiles are difficult to untangle.

Why, given the wealth of fossil material, should the evolution of the reptiles prove so elusive? Is it simply because the major groups have not been properly defined? And, if that is the case, is it because convergence (the evolutionary process whereby unrelated organisms in similar habitats come to superficially resemble each other) is more common than previously recognized? Do the animals look alike because there are unknown structural constraints that mandate their size and shape? The answers to such questions are currently the focus of an intensive research program, which makes it clear that deciphering the past is not an easy task. One comment is certain: the exquisitely preserved San Giorgio fossils have at least enabled paleontologists to ask thought-provoking questions about the nature of life and the processes of evolution.



FOSSIL ASSEMBLAGE from San Giorgio includes the thalattosaur* Askeptosaurus (a)*, a marine reptile about 2.5 meters long that fed on fish;* Ticinosuchus ferox (b)*, a fierce terrestrial species about 2.5 meters in length that was closely related to true dinosaurs; and ichthyosaurs such as this adult* Mixosaurus (c)*, a meter-long marine reptile that bears a superficial resemblance to a dolphin. Careful examination reveals that the mixosaur specimen* (d) *is a pregnant female and is carrying several tiny embryos inside her abdomen.



FURTHER READING

A NEW GENUS OF SHARK FROM THE MIDDLE TRIASSIC OF MONTE SAN GIORGIO, SWITZERLAND. O. Rieppel in *Palaeontology,* Vol. 25, No. 2, pages 399–412; 1982.

ALLOMETRY AND HETEROCHRONY IN THE GROWTH OF THE NECK OF TRIASSIC PROLACERTIFORM REPTILES. Karl Tschanz in *Palaeontology,* Vol. 31, No. 4, pages 997–1011; 1988.

A FOSSIL REPTILE EMBRYO FROM THE MIDDLE TRIASSIC OF THE ALPS. P. Martin Sander in *Science,* Vol. 239, No. 4841, pages 780–783; 1988.

VERTEBRATE PALEONTOLOGY AND EVOLUTION. Robert L. Carroll. W. H. Freeman and Company, 1988.

The Finance of Fission

Wall Street is reluctant to back a resurgence of nuclear power

Nuclear blues, purifying by sunlight, abandoning silicon, the national debt

On the one hand, events seem to be conspiring in favor of nuclear power. During last summer's heat wave, utilities in some regions had to lower voltages below normal to keep all the air conditioners running. The brownouts highlighted the fact that, after years of little or no growth, demand for electricity in the U.S. has crept up by about 5 percent over the past two years. Meanwhile, besieged with accounts of acid rain, the greenhouse effect and other threats to the environment caused by fossil fuels, the public has apparently come to view them as greater evils than nuclear reactors. Shortly after the *Exxon Valdez* spilled 250,000 barrels of oil off the coast of Alaska, a Yankelovich poll found that the public preferred nuclear power to either oil or coal "as the energy source the U.S. should rely on most for its increased energy needs." Little wonder, then, that politicians and pundits who once shunned fission energy are now calling for its revival. As an editorial in the *New Republic* put it: "Everyone, even Pete Seeger, is going to have to take a second look at nuclear power."

On the other hand, Pete Seeger probably could not finance the construction of an atomic reactor, and those institutions that could—namely utilities and the firms that supply them with capital—still see nuclear power as a risky investment. Very risky. "Right now, no utility would touch it with a 10-foot pole," says Herbert Jaffe of the Association of Investors in New York Utilities. As for the investment community itself, it "would not look favorably on any utility that declared its intention to build a reactor at this time," says Carl H. Seligson of Kidder, Peabody and Company. That assessment will only change, he notes, in the event of a "100-percent revamping of the entire process" whereby nuclear reactors are built and regulated.

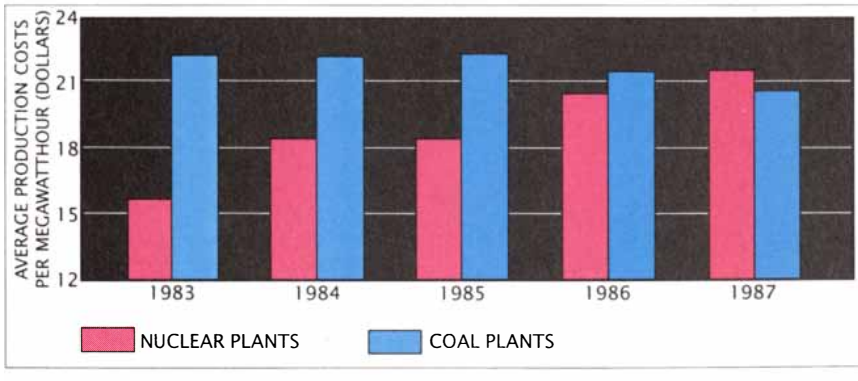
The U.S. already has 110 operating nuclear-power plants, which generate about 20 percent of the nation's electric energy. Only coal-fired plants contribute a greater share (57 percent). But no new reactors have been ordered in more than a decade, and 108 orders have been canceled. The sluggish growth in demand for electricity has been one reason: in recent years utilities have become reluctant to invest in any type of large generator. Another reason is the economic record of the nuclear-power industry. Charles Komanoff of Komanoff Energy Associates, a consulting firm in New York City, observes that in the past 20 years the real costs of building nuclear plants have grown sixfold to almost $4,000 per kilowatt, several times the cost of a typical fossil-fuel plant.

The chief advantage of reactors has traditionally been their low operating cost; unlike fossil-fuel plants they need to be refueled only once every 18 months or so. Yet operating costs have also been steadily rising. In 1987, for the first time, the average coal plant in the U.S. generated power more cheaply than the average nuclear plant, according to the Utility Data Institute.

Utilities contend the Nuclear Regulatory Commission (NRC) is largely to blame for nuclear power's dismal economic record. After the Three Mile Island accident in 1979, the NRC completely rewrote its safety standards, which forced many utilities to retrofit both operating reactors and ones under construction at great expense. Inability to fulfill the requirements in a timely manner has brought financial ruin to several utilities, particularly those trying to gain an operating license for a new plant. (In the case of the ill-starred Shoreham and Seabrook plants, local governments' refusal to participate in emergency evacuation plans mandated by the NRC was the final hurdle.) To reduce such financial risks, utilities have for years sought what they call "one-stop licensing" from the NRC, which would ensure that a utility could operate a plant if it conformed to the original, licensed plan. Recently the NRC proposed such a rule change, but antinuclear groups (yes, a few are still out there) have vowed to fight the proposal in court.

The NRC is only the first stop; after a utility begins operating a plant, it must get the state public service commission to approve a rate hike to help pay for the construction. The commission, which represents the interests of consumers, may reject a proposed rate hike if it decides that a utility spent too much building a plant or that the extra power is no longer needed—and some commissions have done just that. Utilities contend that without a state guarantee of some return on their investment they cannot commit themselves to building a new plant.

The burden of changing the regulatory climate, however, rests with managers of current reactors, suggests Paul C. Parshley of Shearson Lehman Hutton. "They need a better safety record," he says. In spite of steady improvements since the Three Mile Island accident, reactors in the U.S. are still plagued by so many unplanned



COST OF COAL-GENERATED POWER in the U.S. fell below the cost of nuclear power in 1987, even discounting capital costs, according to the Utility Data Institute.

shutdowns that they generate on the average only about two thirds of their capacity. (A recent study by the Massachusetts Institute of Technology showed that reactors in at least five other countries have a better record.) Parshley notes that unplanned outages can cost a utility in several ways: by forcing it to buy replacement power at a higher price, by damaging its credit rating and so driving its interest payments up, and by damaging its reputation in the eyes of the public and of government officials. "Once you're in the penalty box, you're held to a higher standard," he says.

As a way around these regulatory and safety problems, some analysts have advocated the development of a standardized, "inherently safe" reactor. Standardization would supposedly cut operating costs and streamline the regulatory process; utilities might also benefit more from one another's experience than they do now. Senator Timothy E. Wirth has backed the safe-reactor concept in a $4.5-billion bill he has proposed to combat the greenhouse effect; the bill would provide the Department of Energy with $500 million over the next three years to study designs with "passive" safety features, such as a cooling system that relies on gravity rather than pumps to flood an overheated reactor.

Such advanced reactors might be ready for commercial power generation in a decade or so, according to William W. Fish of Donaldson Lufkin & Jenrette Securities. At that point, he says, if the regulatory environment has improved and such critical issues as the disposal of radioactive waste have been resolved, nuclear power in the U.S. might undergo a rebirth. But Fish emphasizes that no utility is likely to invest in a fission reactor if a cheaper source of energy becomes available. "You tell me about the ramifications of cold nuclear fusion," he says, "and I'll tell you about the future of nuclear power." —*John Horgan*

Mr. Clean

Sunlight can destroy dangerous chemicals

Most efforts to harness the free energy that arrives every day as sunshine have focused on generating electric power. In a joint program researchers at the Solar Energy Research Institute (SERI) laboratory in Golden, Colo., and at Sandia National Laboratories in Albuquerque are developing ways to harness the sun's power for a different purpose: to destroy toxic chemicals in industrial waste or groundwater.

***TROUGH-SHAPED MIRRORS** focus sunlight on tubes containing contaminated water. The ultraviolet energy in sunlight initiates a reaction that destroys organic contaminants in the water. Photograph courtesy of Sandia National Laboratories.*

At Sandia, Craig E. Tyner is testing a process that purifies water contaminated with organic chemicals such as trichloroethylene. The water is mixed with titanium dioxide, which serves as a catalyst, and is pumped through Pyrex tubes at the focus of trough-shaped mirrors. The ultraviolet energy in the sunlight generates reactive hydroxyl radicals and peroxide ions that convert a third of the contaminants into benign substances during a 20-second pass over a mirror 120 feet long and seven feet wide. The titanium dioxide is easily filtered out for reuse.

Passing the polluted water through the trough a few times can reduce contaminants present in parts-per-million quantities to levels of parts-per-billion, according to Tyner, who is undertaking larger-scale experiments with groups of mirrors. Tyner envisions unmanned solar detoxification plants that would turn themselves off when the sun is obscured. He believes the technique might be able to purify groundwater near a toxic-waste dump, for example, for as little as one or two dollars per thousand gallons.

Even at that figure, Tyner says, the solar-reaction process is unlikely to compete economically with the commonly used decontamination procedure of air stripping, in which air blown through the contaminated water carries off the organic chemicals. Unlike air stripping, however, the Sandia technique destroys the pollutants rather than releasing them into the atmosphere, and so the technique may become more competitive as pollution standards are tightened.

Still in its early stages is a process being developed by Jim D. Fish of Sandia for converting waste organic chemicals into fuels. The process relies on the sun's heat rather than on its ultraviolet rays. A catalyst (rhodium is under study) is supported in a chamber at the focus of a dish-shaped reflector. The focusing yields temperatures between 800 and 1,000 degrees Celsius, high enough for the catalyst to convert organic chemicals that are pumped through the chamber into such gases as carbon monoxide and hydrogen. The gases can then be made to react, producing methanol for fuel. The process is being tested with methane as a raw material, but ultimately Fish hopes to make fuels out of chlorinated hydrocarbons, which are a menace as pollutants. The waste chlorine could be converted into hydrochloric acid, which is a raw material for many industrial processes.

John P. Thornton of SERI, who is leading the joint effort, has demonstrated another solar-chemistry technique, one that needs no catalyst. In a solar furnace at the White Sands missile range in New Mexico, he has focused light at an intensity of 300 suns on a quartz vessel containing a form of dioxin. The sunlight quickly breaks down 99.9999 percent of the dioxin into relatively harmless compounds. That performance meets the Environmental Protection Agency's standards for the commercial incinerators that are now used to destroy dioxin, some forms of which are highly toxic. According to Thornton, the ultraviolet

energy in sunlight ensures that dioxin is effectively destroyed at temperatures as low as 750 degrees C, hundreds of degrees less than must be attained in an incinerator.

Thornton's scheme is farther from realization than the Sandia projects: to date, he has tested his novel furnace only on milligram quantities of dioxin. Still, he estimates that the first solar-chemistry technologies might be ready for commercialization within five years. *—Tim Beardsley*

So Long, Silicon

U.S. companies slip out of the silicon-wafer business

Silicon-wafer manufacturing is not an industry that often makes headlines. Worldwide sales are modest—about $2.2 billion last year. The technology is mature, having been developed primarily at the Bell Telephone Laboratories in the 1950's. But when Monsanto sold its wafer division to a West German company this spring, wafer making joined the ranks of industries pioneered and then abandoned by U.S. companies.

Wafers of single-crystal silicon—made by melting chunks of polycrystalline silicon and seeding the melt to grow single crystals—are the bedrock of the more profitable microelectronics industry. Although a handful of companies, including AT&T, IBM and Texas Instruments, make their own silicon wafers, most chip manufacturers buy wafers from vendors, etch integrated-circuit patterns onto the silicon and sell the chips.

But the business has proved too cyclical for most U.S. wafer vendors' balance sheets, says Robert L. Kelley, formerly the general counsel for the Monsanto Electronic Materials Company (MEMC).MEMC'sexperienceswereprobably typical; it lost roughly $270 million from 1983 to 1987 and posted an $11-million operating gain last year. In large part because of such vagaries, the wafer division fell out of step with the parent company's strategic plans.

Silicon wafers do fit into the plans of the Department of Defense, however. A recent Defense report listed the technology as critical to the long-term superiority of U.S. weapons systems. As Monsanto prepared to sell MEMC to Heuls AG, a subsidiary of a West German conglomerate, the Government called for a review of the national security implications of the sale.

"Try to imagine the sale of the last U.S.-owned steel mill," says William C. O'Mara, an industry consultant based in Palo Alto, Calif. Without a U.S.-owned supplier, microchip makers might not always receive wafers on time, argued SEMATECH, a chip-manufacturing research consortium, and SEMI, a related consortium of equipment manufacturers.

The Government review, however, came too late to preserve the industry. Other U.S. silicon-wafer makers had sold out after suffering huge losses in 1985. Monsanto had frozen funding for capital expenditures for MEMC in 1986. Even finding a buyer for the division was not simple; the only serious contenders were Heuls and several U.S. financial firms that specialize in leveraged buy-outs. On February 7 President Bush agreed to let the sale of MEMC to the West German company go through.

Oddly, there are still U.S. companies making polycrystalline silicon, the raw material for the wafers. But these firms have also been buffeted by ups and downs in the demand for semiconductors and are looking overseas for funds. In 1986 Japan's second largest steelmaker, Nippon Kokan K. K., entered the industry by buying the Great Western Silicon Company from General Electric. Nippon Kokan is also considering entering other segments of the microchip business, including wafer manufacturing and integrated-circuit design. "If we can succeed in all parts of the business," says Susumu Terao, a manager at Great Western Silicon, "maybe we can make a big profit." *—Elizabeth Corcoran*

THE ANALYTICAL ECONOMIST

Is the U.S. national debt a problem?

On the corner of Forty-Second Street and Sixth Avenue in Manhattan, not far from the tawdry lights of Times Square, is a blinking sign. "Our National Debt: $2,758,084,675," it declares. At least that was the size of it in mid-April. Since the figure grows by $8,000 every second, the numbers have long since changed.

The sign may seem a gloomy presage of economic decay, but economists are far from unanimous in their assessment of the debt's significance. Some blame the debt for a host of economic ills, including declining savings and investment and increasing interest rates, and for raising the specters of inflation and recession. Others argue that household savings can balance the Government's "dissavings" or that the Government's bookkeeping exaggerates the size of the debt. Most agree, however, that arguments over the debt point to more fundamental questions about the Government's role in the economy and who should pay for the costs of that role.

The debt first became a public policy issue during the Great Depression. Relying largely on ideas advanced by John Maynard Keynes, the Roosevelt Administration tried to stimulate the economy by spending beyond its revenues in order to create jobs. Keynes had postulated that consumers decide to spend or save on the basis of their income and that businesses decide to spend or invest depending on how strong they expect consumer demand will be. A government, then, could pump up the economy by cutting taxes and increasing spending or could curb the economy by raising taxes; deficits incurred when stimulating the economy should not cause alarm.

By the end of World War II, Keynesian economics had won wide acceptance. "Annual budget-balancing is, thus, both difficult in practice and unsound in principle," a group of economists told Congress in 1949. "But one great merit it does have: it provides a yardstick by which [to measure] each activity of the government."

But by the mid-1970's the debt began gaining national notoriety in the wake of analyses by Martin S. Feldstein, an economist at Harvard University (and later President Reagan's chief economist), that linked deficits to declining savings. Other economists began criticizing the debt for competing with the private sector for consumer savings and thereby driving up interest rates. Higher interest rates make it more costly for businesses to borrow money for capital investment—leading to lower investment and economic stagnation. The Government might print more money to pay off some of the debt—but that would fuel inflation. The debt debate became shrill as it moved into the political arena, and soon continuing deficits looked like a short route to disaster.

Yet for the past seven years the U.S. economy has belied the doomsayers. Since the 1981–1982 recession, economic statistics have recorded a mounting debt but at the same time

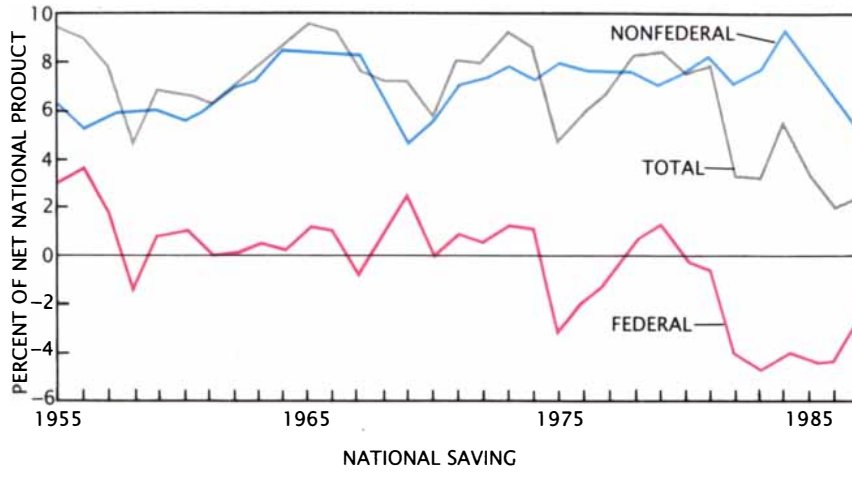
low inflation and high employment. The stock market enjoyed record growth until the October 1987 crash; it has since recovered. As a percentage of the gross national product, the U.S. debt is somewhat higher than the debt of Japan, West Germany and France but lower than that of Great Britain, Canada and Italy. Net foreign investment, traditionally a minuscule part of U.S. accounts, has risen to more than 3 percent of the gross national product, relieving pressure on interest rates. The economy's health has perplexed many economists and added weight to models that do not consider the $2.7-trillion debt very important.

One debt dissenter is Robert J. Barro at Harvard, who draws on the work of 19th-century English economist David Ricardo. Barro argues that consumers are rational: when they see a government running up large budget deficits, they realize they will eventually have to bear the costs. They increase their savings to balance the spending.

Of course, few applaud the present U.S. rate of personal savings, which fell to about 3 percent of disposable personal income in 1987. (Personal savings averaged more than 7 percent from 1950 to 1980.) Barro prefers a different measure of savings: the net worth of households rather than just their savings accounts. He would include such items as consumer durables (automobiles, for instance) and capital gains on stocks and real estate. In this way Barro calculates that the real value of households has averaged 10 to 11 percent of the gross national product during the past four years, a figure that makes him comfortable with the size of the debt.

Other economists argue that the size of the debt is exaggerated by what the Government includes or leaves out of its tallies. Robert Eisner of Northwestern University (the 1988 president of the American Economic Association) argues that the present accounts poorly reflect U.S. economic activity and value. Eisner and others would, for instance, adjust each annual deficit for inflation in order to measure the debt's real value. They would also add state and local governments' budget surpluses to the federal accounts. (State and local governments have saved about 9 percent of their annual receipts since 1980.)

Eisner is also troubled by the inability of national accounts to distinguish between Government "consumption" and "investment." For example, federal expenditures for research, education and new roads all fall into the category of purchasing goods and services. Eisner and others calculate that U.S. net investment may be several times the level portrayed in the Government books. Because critics of the debt charge that it is absorbing funds that could otherwise finance capital investment, Eisner suggests separating the Government's "investment" spending from its debt tab.



CHANGES IN SAVINGS (with adjustments for inflation) were mapped as a percent of net national product by Edward M. Gramlich at the University of Michigan. His calculations, based on numbers from national accounts, show federal savings* (red) *have slipped the most; since 1984, state- and local-government and private savings* (blue) *have also dropped. As a result, total national savings* (gray) *are low. Courtesy of the* Journal of Economic Perspectives, *Spring 1989.

Laurence J. Kotlikoff of Boston University would abandon the present federal accounts altogether. He argues that distinguishing "revenues" (which the Government earns with taxes) from "loans" (which represent the debt) only distracts attention from the question of which generations will bear the heaviest cost for Government spending. Kotlikoff points out that a tax break today encourages spending, particularly by older citizens, and is eventually paid for by their children. Because the size of the debt does not reveal who pays the cost of public programs, Kotlikoff suggests that the Government instead try to estimate how its policies will affect different generations' lifetime income and then balance the costs appropriately.

Kotlikoff's critics argue that his approach is unworkable from a practical standpoint, but other economists echo his concerns. One is Michael J. Boskin, formerly of Stanford University and now economic advisor to President Bush. Moreover, like Eisner, Boskin has his own program for revising the national accounts: he would include state and local budgets and the value of such federal capital assets as mineral and oil rights. Even though Boskin's calculations trim the level of the debt, he still sees "substantial analytical and empirical findings that the debt discourages investment."

Many economists who believe that the danger of the debt is overstated still worry about where the U.S. will find savings for future investment, particularly for public goods such as education and infrastructure. Most agree that the debt represents some amount of dissavings, although they argue about how much. Household savings have also slipped; even Barro concedes that household wealth falls short of his expectations.

In recommending fiscal strategies, these economists nevertheless part company with those who are deeply concerned about the debt. The depressed savings rate does not mean the Government must slash the debt by cutting spending across the board, they say; instead it should look afresh at those expenditures that are strictly "consumption" and those more appropriately called "investments" and should pay more attention to the long-term effects of fiscal policy.

Sorting out consumption from investment is fraught with difficulties, however, and different versions of the national accounts reflect their authors' biases. Although economics can help distinguish between consumption and investment, it cannot dictate precise categories. Federal-spending choices ultimately rest with politicians and the citizens who elect them. The current debt figure offers little insight on which to base such decisions.

—Elizabeth Corcoran and Paul Wallich

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.

Whether it's wide area broadband, speech recognition, or exciting new probes into expert system computer logic, NYNEX scientists are bringing technology to fruition.

The NYNEX family of companies and the Science & Technology Center are working together to provide leading edge answers to our customers in the areas of advanced computer networks, software, mobile communications, local area networks and telecommunication networks.

To find out why the answer is NYNEX, and how our scientists and technology can work with your company, call us at 1 800 535-1535.

Our technology will help you tower over the competition.

Need to communicate? Need to compute? The answer is

NYNEX

The Fourier Transform

DNA's double helix, the sunspot cycle and the sawtooth signals of electronics can be reduced mathematically to a series of undulating curves. This idea underlies a powerful analytical tool

by Ronald N. Bracewell

To calculate a transform, just listen. The ear automatically performs the calculation, which the intellect can execute only after years of mathematical education. The ear formulates a transform by converting sound—the waves of pressure traveling through time and the atmosphere—into a spectrum, a description of the sound as a series of volumes at distinct pitches. The brain turns this information into perceived sound.

Similar operations can be done by mathematical methods on sound waves or virtually any other fluctuating phenomenon, from light waves to ocean tides to solar cycles. These mathematical tools can decompose functions representing such fluctuations into a set of sinusoidal components—undulating curves that vary from a maximum to a minimum and back, much like the heights of ocean waves. The Fourier transform is a function that describes the amplitude and phase of each sinusoid, which corresponds to a specific frequency. (Amplitude describes the height of the sinusoid; phase specifies the starting point in the sinusoid's cycle.)

The Fourier transform has become a powerful tool in diverse fields of science. In some cases, the Fourier transform can provide a means of solving unwieldy equations that describe dynamic responses to electricity, heat or light. In other cases, it can identify the regular contributions to a fluctuating signal, thereby helping to make sense of observations in astronomy, medicine and chemistry.

RONALD N. BRACEWELL has been on the faculty of the electrical engineering department at Stanford University since 1955. He was educated at the University of Sydney and the Cavendish Laboratory in Cambridge, England, where he got his Ph.D. His research interests have ranged over microwave radar, ionospheric physics and radio astronomy. At Stanford he is a member of the Space, Telecommunications and Radioscience Laboratory and is professor of computer science, by courtesy.

The world first learned about the technique from the mathematician for whom the transform is named, Baron Jean-Baptiste-Joseph Fourier. Fourier was not merely interested in heat; he was obsessed by it. He kept his home in Grenoble so uncomfortably hot that visitors often complained. At the same time he would cloak himself in heavy coats. Perhaps it was the lure of a warm climate that in 1798 drew Fourier to join the retinue of 165 savants that accompanied Napoleon's expedition to Egypt.

While Napoleon was fighting Syrians in Palestine, repelling the Turks from Egypt and hunting the Mameluke chief, Murad Bey, the French scientists undertook ambitious studies in geography, archaeology, medicine, agriculture and natural history. Fourier was appointed secretary of a scientific body known as the Institute of Egypt. He discharged administrative duties with such competence that he received many diplomatic assignments. Yet he was still able to conduct intensive research on Egyptian antiquities and contemplate a theory about the roots of algebraic equations.

Shortly before the French were driven from Egypt in 1801, Fourier and his colleagues set sail for France. The commander of the British fleet, Admiral Sir Sidney Smith, promptly seized their ship along with its cargo of Egyptian documents and relics. In the honorable spirit of the time, Smith put the scientists ashore unharmed in Alexandria. The English commander eventually traveled to Paris to return the confiscated material—except for the Rosetta stone (the key to Egyptian hieroglyphics), which stands today in the British Museum as a monument to Napoleon's military defeat and his contribution to Egyptology.

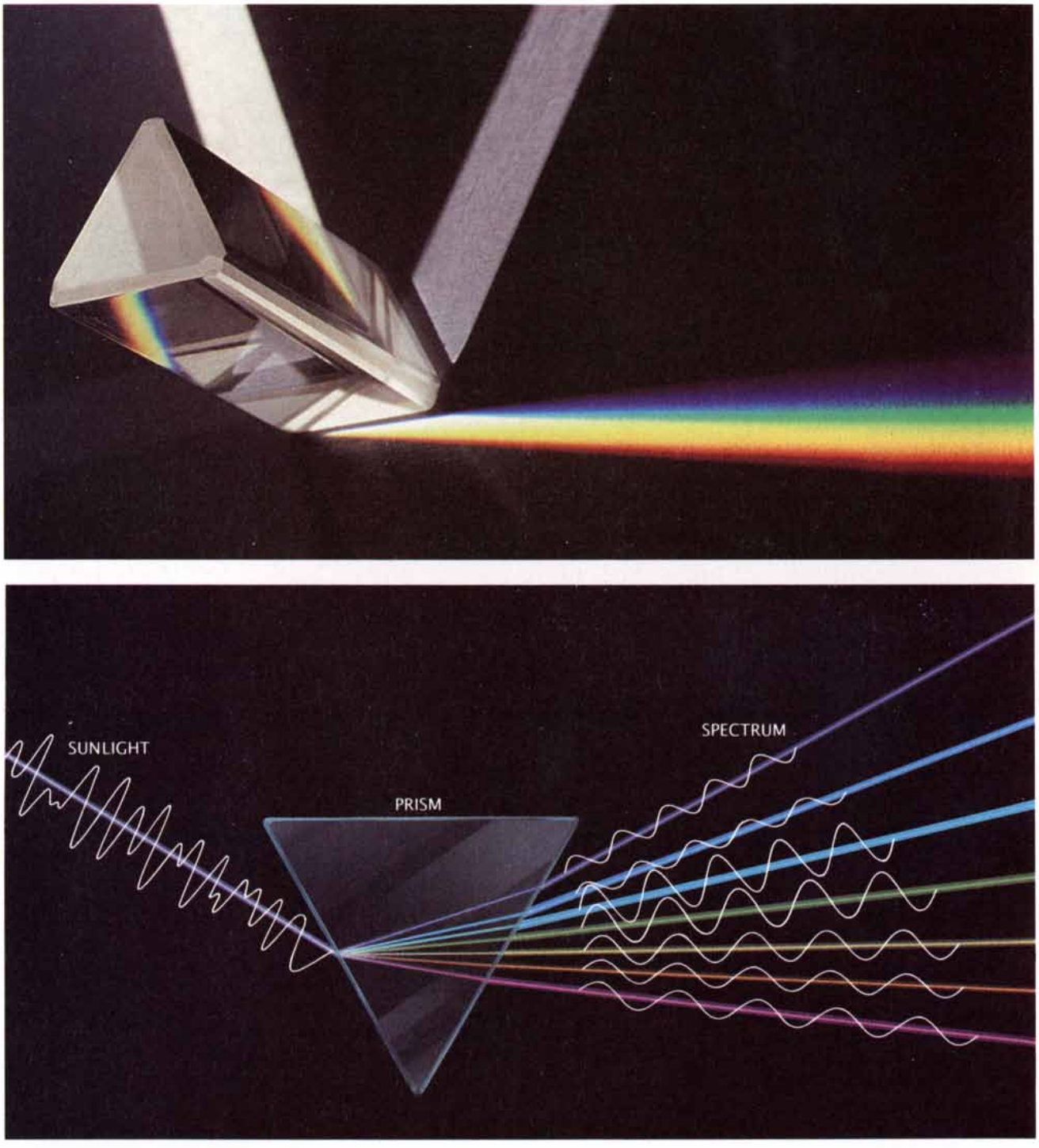
Returning to France relatively unscathed, Fourier focused on mathematical matters as professor of analysis at the Polytechnic School, but in 1802 he again entered Napoleon's service. Fourier became the prefect of the Isère department. While attempting to repair the disruptions remaining from the Revolution of 1789, he built the French section of the road to Turin and drained 80,000 square kilometers of malarial swamp. During this time he derived an equation that described the conduction of heat in solid bodies. By 1807 Fourier had invented a method for solving the equation: the Fourier transform.

Fourier applied his mathematical technique to explain many instances of heat conduction. A particularly instructive example that avoids computational complications is the flow of heat around an anchor ring—an iron ring that attaches a ship's anchor to its chain—that has been thrust halfway into a fire. When part of the circumference becomes red hot, the ring is withdrawn. Before much heat is lost to the air, the ring is buried in fine, insulating sand, and the temperature is measured around the outer curve [*see illustration on page 88*].

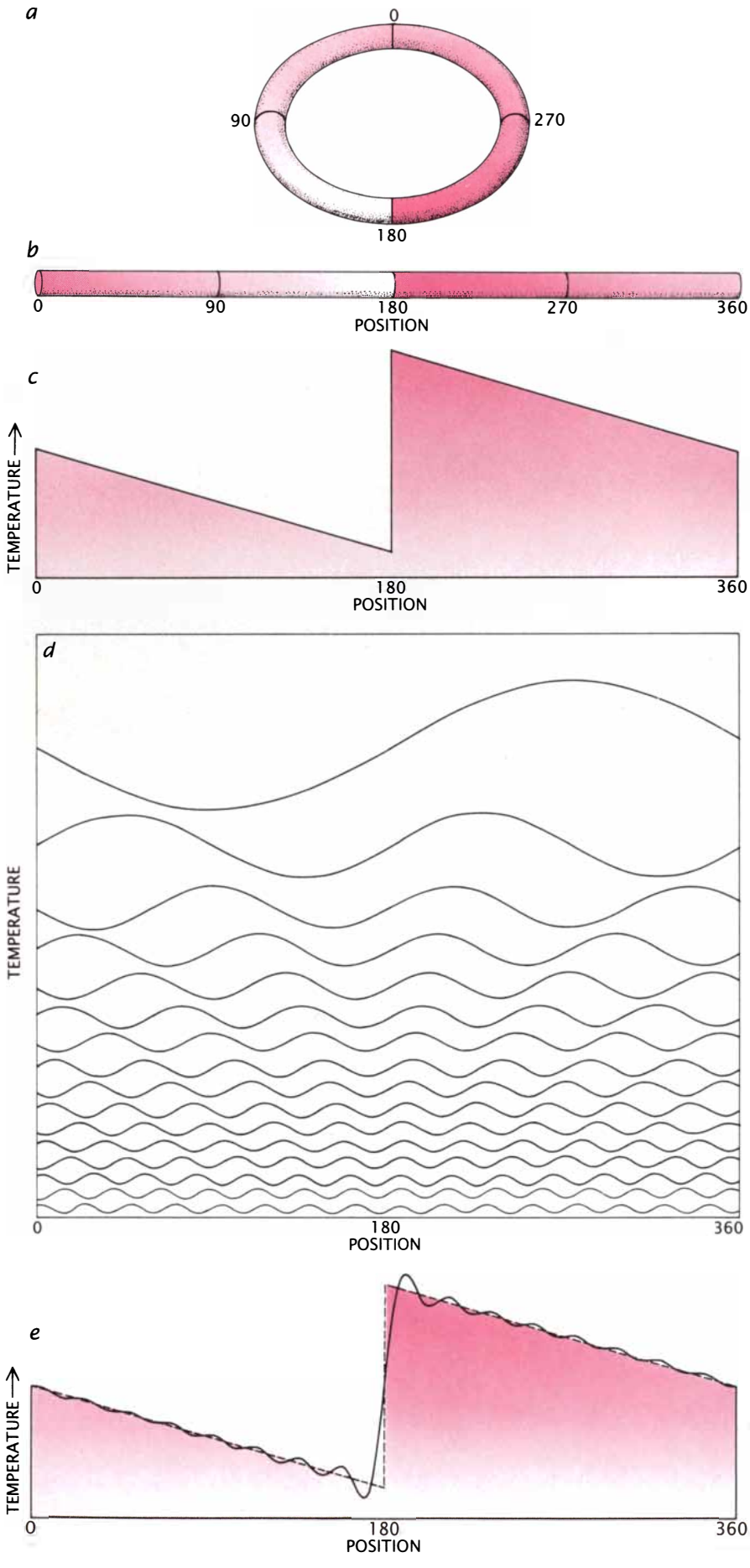
Initially the temperature distribution is irregular: part of the ring is uniformly cool, and part is uniformly hot; in between the temperature abruptly shifts. As heat is conducted from the hot region to the cool region, however, the distribution begins to smooth out. Soon the temperature distribution of heat around the ring reaches a sinusoidal form: a plot of the temperature rises and falls evenly, like an S curve, in exactly the way sine and cosine functions vary. The sinusoid gradually flattens until the whole ring arrives at a constant temperature.

Fourier proposed that the initial, irregular distribution could be broken down into many simple sinusoids

that had their own maximum temperature and phase, that is, relative position around the ring. Furthermore, each sinusoidal component varied from a maximum to a minimum and back an integral number of times in a single rotation around the ring. The one-cycle variation became known as the fundamental harmonic, whereas variations with two, three or more cycles in a single rotation became the second, third and higher harmonics. The mathematical function that describes the maximum temperature and position, or phase, for each of the harmonics is the Fourier transform of the temperature distribution. Fourier had traded a single distribution that was difficult to describe mathematically for a more manageable series of full-period sine and cosine functions



SUNBEAM resolved into a spectrum provides a physical analogy for mathematical transforms (*top*). The sunlight entering the prism varies in strength from moment to moment (*bottom*). The light leaving the prism has been separated in space into pure colors, or frequencies. The intensity of each color implies an amplitude at each frequency. Thus, a function of strength versus time has been transformed into a function of amplitude versus frequency. The Fourier transform can represent a time-varying signal as a function of frequency and amplitude, but the transform also provides information about phase.



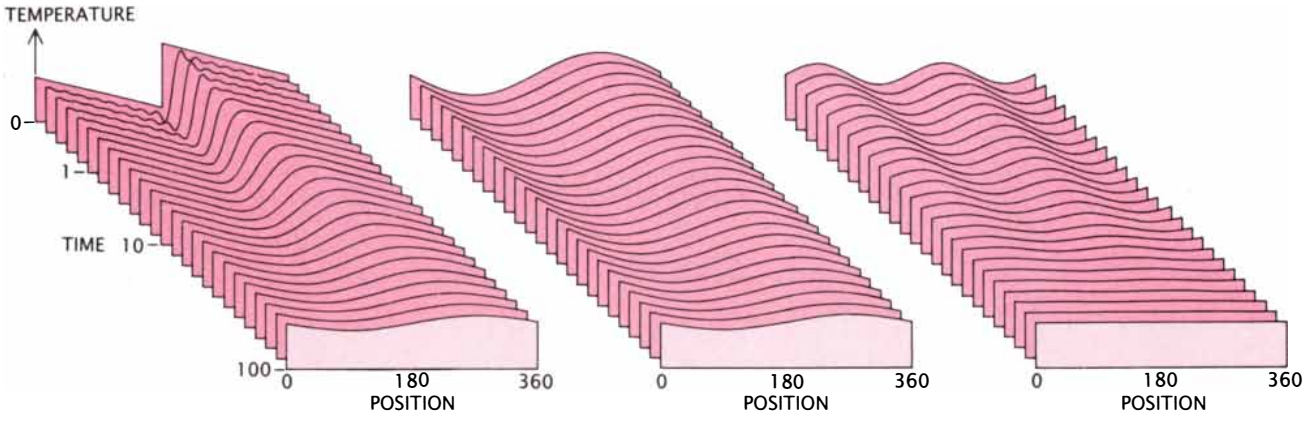
that when added together would make up the original distribution.

In applying this analysis to the conduction of heat around the ring, Fourier reasoned that the greater the number of periods of a sinusoidal component, the more rapidly it will decay. One can follow his reasoning by examining the relation between the fundamental and the second harmonic of the temperature distribution. The temperature of the second harmonic varies from hot to cool twice around the circumference of the ring, whereas the fundamental varies only once. Therefore, the distance that the heat must travel from hot peak to cool trough is only half as far for the second harmonic as it is for the fundamental. Furthermore, the temperature gradient in the second harmonic is twice as steep as it is in the fundamental variation. Because twice the heat flow occupies half the distance, the second harmonic will die out four times faster than will the fundamental.

Higher harmonics will decay even more rapidly. Hence, it is only a single sinusoidal distribution of the fundamental variation that persists as the temperature of the ring approaches equilibrium. Fourier believed that the evolution over time of any initial heat distribution could be computed by his technique.

Fourier's analysis challenged the mathematical theories to which his contemporaries adamantly adhered. In the early 19th century, many extraordinary Parisian mathematicians, including Lagrange, Laplace, Legendre, Biot and Poisson, could not accept Fourier's claim that any initial temperature distribution could be decomposed into a simple arithmetic sum that consisted of a fundamental variation and its higher-frequency harmonics. Leonhard Euler also found fault with Fourier's ideas, although he had already proposed that some

TEMPERATURE OF AN IRON RING was one of the first phenomena analyzed by Fourier's technique. One distribution of heat around a ring is shown (*a*); brighter color represents hotter areas. To begin the analysis, the ring is "uncoiled" (*b*), and the temperature is measured at every point, yielding a temperature distribution around the circumference (*c*). Then the temperature distribution is decomposed into many sinusoidal curves having one, two, three or more cycles (*d*). When 16 of the curves are simply added together (*solid line in e*), they yield a good approximation of the original temperature distribution (*broken line in e*).



CONDUCTION OF HEAT through an iron ring causes the temperature distribution to change over time (*left*). Just as the temperature distribution at any instant can be described as a series of sinusoidal curves, the evolution of a temperature distribution over time can be described in terms of changes in the sinusoids. The one-cycle distribution, or first harmonic (*middle*), and the two-cycle distribution, or second harmonic (*right*), are shown. Fourier determined that the second harmonic will decay four times faster than the first harmonic and higher harmonics will decay even faster. Because the first harmonic persists the longest, the overall temperature distribution approaches the sinusoidal shape of the first harmonic.

functions could be represented as a sum of sine functions. And so when Fourier made this claim at a meeting of the French Academy of Sciences, Lagrange stood up and held it to be impossible.

Even under these circumstances the Academy could not ignore the significance of Fourier's results, and it awarded him a prize for his mathematical theory of the laws of heat propagation and his comparison of the results of his theory with precise experiments. The award was announced, however, with the following caveat: "The novelty of the subject, together with its importance, has decided us to award the prize, while nevertheless observing that the manner in which the author arrives at his equations is not without difficulties, and that his analysis for integrating them still leaves something to be desired both as to generality and even as to rigor."

The great uneasiness with which Fourier's colleagues regarded his work caused its publication to be delayed until 1815. In fact, it was not completely described until the 1822 publication of his book, *The Analytical Theory of Heat*.

Objections to Fourier's approach focused on the proposition that an apparently discontinuous function could be represented by a sum of sinusoidal functions, which are continuous. Discontinuous functions describe broken curves or lines. For instance, a function called the Heaviside step function is zero on the left and jumps to one on the right. (Such a function can describe the flow of current when a switch is turned on.) Fourier's contemporaries had never seen a discontinuous function described as resulting from a combination of ordinary, continuous functions, such as linear, quadratic, exponential and sinusoidal functions. If Fourier was correct, however, a sum of an infinite number of sinusoids would converge to represent accurately a function with jumps, even with many jumps. At the time this seemed patently absurd.

In spite of these objections many workers, including the mathematician Sophie Germain and the engineer Claude Navier, began extending Fourier's work beyond the field of heat analysis. Yet mathematicians continued to be plagued by the question of whether a sum of sinusoidal functions would converge to represent a discontinuous function accurately.

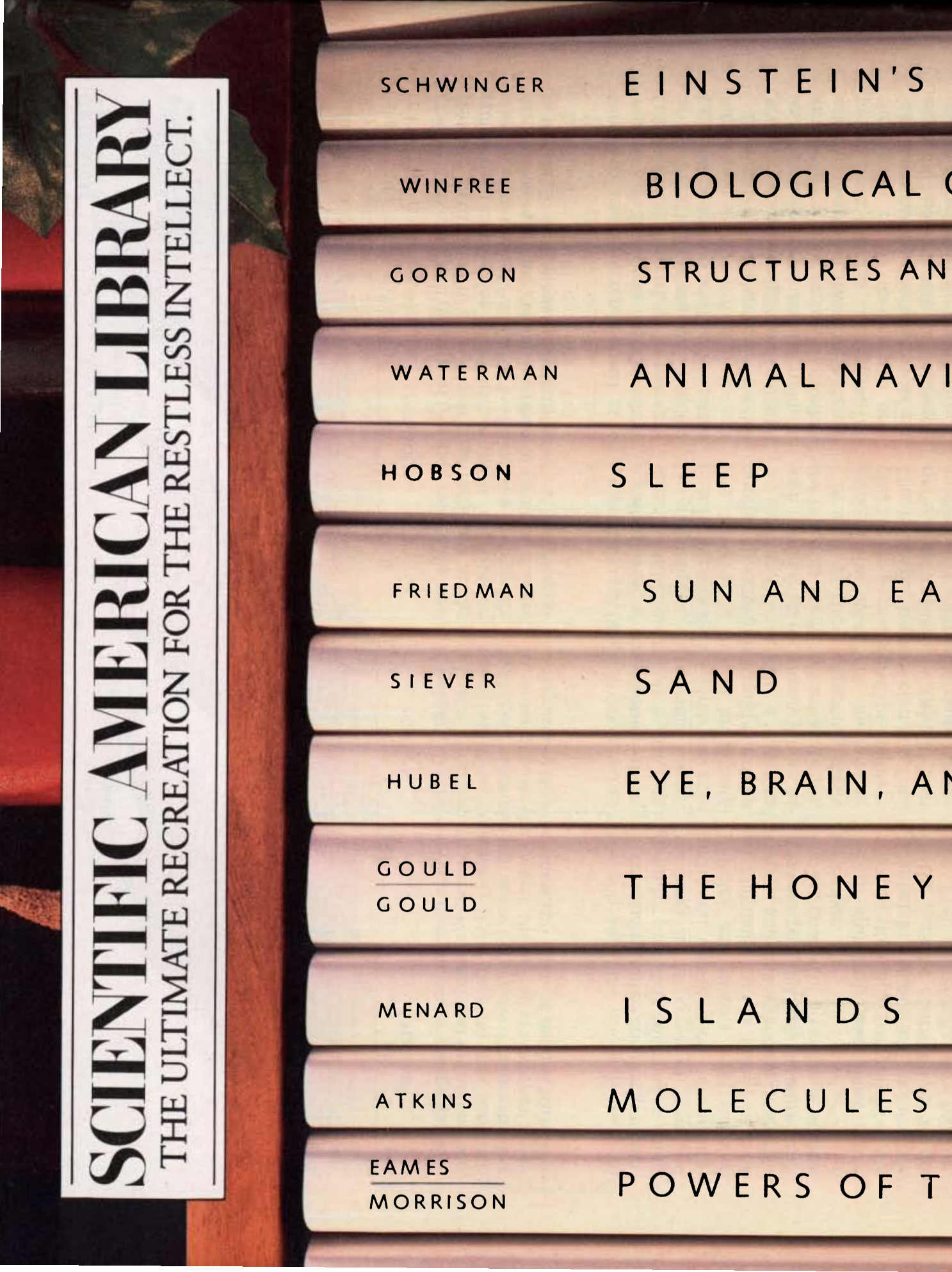
The question of convergence arises whenever an infinite series of numbers is to be added up. Consider the classic example: will you ever arrive at a wall if with each step you travel half of the remaining distance? The first step will bring your toe to the halfway mark, the second, three quarters of the way, and at the end of the fifth step you are almost 97 percent of the way there. Clearly this is almost as good as reaching the wall, but no matter how many steps you take, you will never quite reach it. You could prove mathematically, however, that you would ultimately get closer to the wall than any distance nominated in advance. (The demonstration is equivalent to showing that the sum of a half, a fourth, an eighth, a 16th and so on approaches one.)

The question of the convergence of Fourier series emerged again late in the 19th century in efforts to predict the ebb and flow of the tides. Lord Kelvin had invented an analogue computer for providing information about the tides to the crews of merchant and naval vessels. First sets of amplitudes and phases were calculated manually from a record of tidal heights and corresponding times that had been painstakingly measured during the course of a year in a particular harbor.

Each amplitude and phase represented a sinusoidal component of the tidal-height function and revealed one of the periodic contributions to the tide. Then the results were fed into Lord Kelvin's computer, which synthesized a curve predicting the heights of the tide for the next year. Tidal curves were soon produced for ports all over the world.

It seemed obvious that a tide-predicting machine with more parts could process more amplitudes and phases and thus would make better predictions. This turned out not to be completely true if the mathematical function to be processed contained a steep jump, that is, it described an essentially discontinuous function.

Suppose such a function was reduced into a small set of amplitudes and phases—that is, just a few Fourier coefficients. The original function can then be reconstructed from the sinusoidal components corresponding to the coefficients, and the error between the original function and the reconstructed function can be measured at each point. The error-finding procedure is repeated, each time computing more coefficients and incorporating them into the reconstruction. In every case, the value of the maximum error does not diminish. On the other hand,


SCIENTIFIC AMERICAN LIBRARY
THE ULTIMATE RECREATION FOR THE RESTLESS INTELLECT.
SCHWINGER
EINSTEIN'S
WINFREE
BIOLOGICAL
GORDON
STRUCTURES AN
WATERMAN
ANIMAL NAVI
HOBSON
SLEEP
FRIEDMAN
SUN AND EA
SIEVER
SAND
HUBEL
EYE, BRAIN, AN
GOULD
GOULD
THE HONEY
MENARD
ISLANDS
ATKINS
MOLECULES
EAMES
MORRISON
POWERS OF T

Join the most exciting book club in science—where the only commitment you make is to reward your own curiosity

This is the Scientific American Library, where the mysteries of the human and natural worlds unfold before your eyes in lavishly illustrated volumes you'll want to read and refer to for years to come.

Here's why 75,000 readers like you have already become members of the Scientific American Library

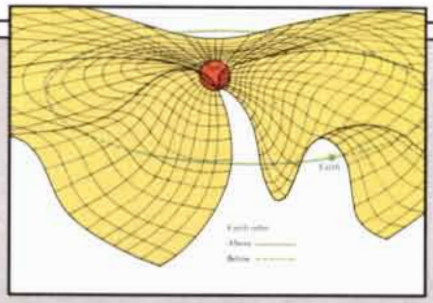
■ **15-day Free Trial**
You will receive a new volume to examine approximately every two months. (Your first volume is shipped immediately via UPS at no extra charge.) You have 15 days to look over each volume. If you decide to keep it, send in your payment. Otherwise, simply return the book in the same shipping carton and we'll immediately credit your account.

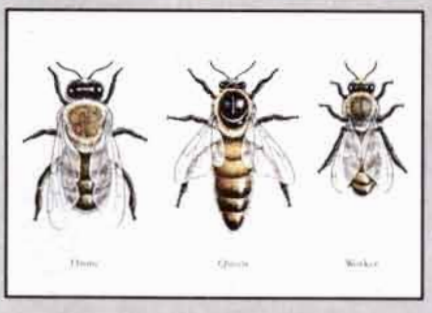
■ **No Prepayment and No Minimum Purchase**
Send no money in advance. You pay only for the volumes you keep. You may choose as many or as few volumes as you like: there is no yearly minimum.

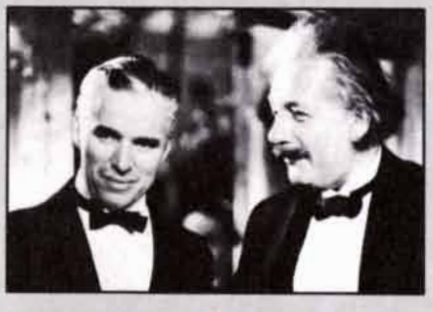
■ **Low Members' Price**
You may purchase any of these handsome volumes for only $24.95—25% below the bookstore price.

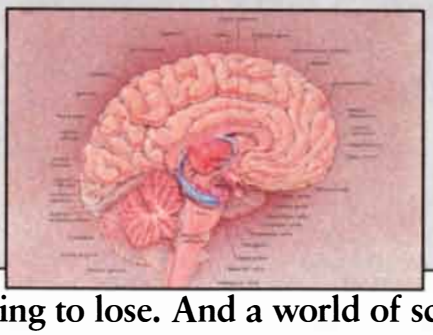
■ **No Strings, No Risk, No Commitment**
You may cancel your obligation-free membership at any time for any reason, without notice, without penalty.

CREDITS (top to bottom)—From Sun and Earth, *the "ballerina skirt" model of the solar-interplanetary current sheet. From* The Science of Structures and Materials, *bubble formation due to the surface tension of a liquid; courtesy Stephen Dalton, NHPA. From* The Honey Bee, *drone, queen, and worker; drawing by John Hatzakis. From* Einstein's Legacy, *Chaplin and Einstein in Hollywood, 1931; courtesy The Granger Collection. From* Drugs and The Brain, *major anatomical divisions of the brain; painting by Bill Andrews.*









A one-of-a-kind book club for those who take science seriously.

Each volume in the Scientific American Library was created exclusively for this series, written by preeminent scientists—including many Nobel laureates—who describe their work in an exciting, accessible way.

Choose your premier volume

■ **Extinction,** Steven M. Stanley. In this geologic detective story, a renowned paleobiologist explores the causes of the mass extinctions that have periodically decimated thousands of thriving species. "Smack up to date scientifically...the best overview of the changing relationship between the physical environment and life.... Enthralling...."
—John Gribbon, *New Scientist*

■ **Molecules,** P. W. Atkins. A distinguished chemist takes us below the surface of visible reality to show how molecules make soap "soapy," determine the taste of barbecue, and give rise to the brilliant colors of fall foliage. "Undoubtedly the most beautiful chemistry book ever written.... It is more than a book of facts; *Molecules* is a work of art."
—John Emsley, *New Scientist*

Join us. You have nothing to lose. And a world of scientific discovery awaits.

To become a member of the Scientific American Library, simply detach and mail the facing reply card.
Or write to: Scientific American Library, P.O. Box 646, Holmes, PA 19043

the error becomes confined to a region that gradually shrinks around the discontinuity, so that ultimately at any given point the error approaches zero. Josiah Willard Gibbs of Yale University confirmed this result theoretically in 1899.

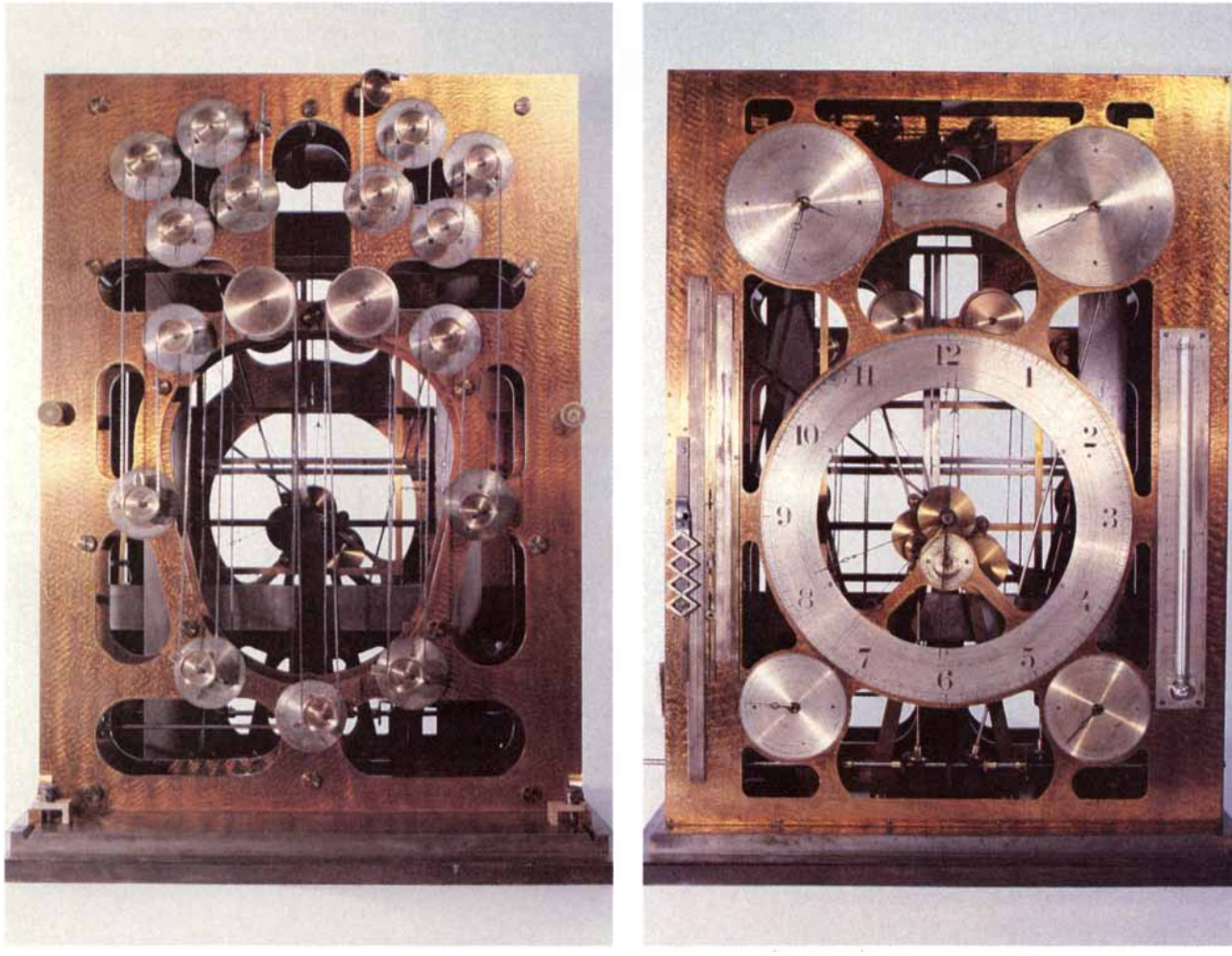
Fourier analysis is still not applicable to unusual functions, such as those possessing an infinite number of infinite jumps in a finite interval. By and large, however, a Fourier series will converge if its original function represents the measurement of a physical quantity.

Vast areas of new mathematics have been developed from investigations of whether the Fourier series of a particular function converges. One example is the theory of generalized functions, which is associated with George F. J. Temple of England, Jan G. Mikusiński of Poland and Laurent Schwartz of France. It established in 1945 a firm basis for the Heaviside step function and the Dirac delta function; the latter describes a unit of area concentrated at a point. The theory enabled the Fourier transform to be applied to solve equations that involved such intuitively accepted concepts as point mass, point charge, magnetic dipoles and the concentration of a load on a beam.

After almost two centuries of development, the theory behind the Fourier transform is firmly established and well understood. As we have seen, Fourier analysis breaks down a function in space or time into sinusoidal components that have varying frequencies, amplitudes and phases. The Fourier transform is a function that represents the amplitude and phase at each frequency. The transform can be derived by two different mathematical methods, one applied when the original function is continuous and the other when it consists of many discrete measurements.

If the function is made from a list of values at discrete intervals, it can be broken down into a series of sinusoidal functions at discrete frequencies, which range from a lowest frequency, the fundamental, through a series of frequencies that are two, three or more times the fundamental. Such a sum of sinusoids is called the Fourier series.

If the original function provides a value for every real number, then it is decomposed into sinusoidal functions at all frequencies, which are combined by means of an operation called the Fourier integral. The transform is neither the series nor the integral. In the case of the discrete function, it is the frequency-dependent list of amplitudes and phases appearing in the Fourier series; in the case of the



FERREL TIDE PREDICTOR, an analogue computer built in the late 19th century, performed Fourier synthesis to forecast the ebb and flow of the tides. Data that were collected on tidal heights at a particular harbor could be reduced by hand calculations into a set of numbers, each one representing a periodic contribution to the tide, such as the gravitational pull of the moon. The numbers for a specific port could then be fed into the Ferrel Tide Predictor by twisting knobs on the back of the machine (*left*). When a time was set on the front of the machine (*right*), the predicted height of the tide could be read off a dial.

unbroken function, it is the function of frequency that results when the Fourier integral is evaluated.

Regardless of the manner in which the transform is derived, it is necessary to specify two numbers at each frequency. These might be the amplitude and phase; however, other number pairs could encode the same information. These values can be expressed as a single complex number. (A complex number is the sum of one real number and another real number multiplied by the square root of negative one.) This representation is very popular because it invites the use of complex algebra. The theory of complex algebra and the Fourier transform have become indispensable in the numerical calculations needed to design electrical circuits, analyze mechanical vibrations and study wave propagation.

Representing an original function by its complex Fourier transform leads to computational advantages. A typical problem is to ascertain the current that flows when a known voltage is applied to a circuit. The direct method involves solving a differential equation that relates the voltage and current functions. The Fourier transforms of the voltage and current function, in contrast, can be related by an equation whose solution is trivial.

Today the study of Fourier transforms consists largely of acquiring techniques for moving freely between functions and their transforms. Analytical methods can be applied to evaluate the Fourier integral and produce the transform. Although these methods may be difficult for ordinary practitioners, many Fourier integrals have been found and are listed in tables of reference. These methods can be supplemented by learning a handful of theorems pertaining to transforms. With the aid of these theorems more or less complicated wave forms can be handled by reduction to simpler components.

Fortunately numerical methods are available for computing Fourier transforms of functions whose forms are based on experimental data or whose Fourier integrals are not easily evaluated and are not found in tables. Before electronic computers, numerical calculation of a transform was rather tedious, because such a large amount of arithmetic had to be performed with paper and pencil. The time required could be reduced somewhat by forms and schedules that guided investigators through the calculations, but the labor involved could still be daunting.

Just how much arithmetic had to be performed depended on the number of data points needed to describe the wave. The number of additions was comparable to the number of points, and the number of multiplications equaled the number of points squared. For example, analyzing a wave specified by 1,000 points taken at regular intervals required on the order of 1,000 additions and one million multiplications.

Such calculations became more feasible as computers and programs were developed to implement new methods of Fourier analysis. One was developed in 1965 by James W. Cooley of IBM's Thomas J. Watson Research Center and John W. Tukey of the Bell Telephone Laboratories in Murray Hill, N.J. Their work led to the development of a program known as the fast Fourier transform.

The fast Fourier transform saves time by decreasing the number of multiplications needed to analyze a curve. At the time, the amount of multiplication was emphasized simply because multiplication was slow with respect to other computer operations, such as addition and fetching and storing data.

The fast Fourier transform divides a curve into a large number of equally spaced samples. The number of multiplications needed to analyze a curve decreases by one half when the number of samples is halved. For example, a 16-sample curve would ordinarily take 16 squared, or 256, multiplications. But suppose the curve was halved into two pieces of eight points each. The number of multiplications needed to analyze each segment is eight squared, or 64. For the two segments the total is 128, or half the number required before.

If halving the given sequence yields a twofold gain, why not continue with the strategy? Continued subdivision leaves eight irreducible pieces of two points each. The Fourier transforms of these two-point pieces can be computed without any multiplications, but multiplication is required in the process of combining the two-point transforms to construct the whole transform. First, eight two-point transforms are combined into four four-point transforms, then into two eight-point transforms and finally into the desired 16-point transform. These three stages that combine the pieces each call for 16 multiplications, and so the total number of multiplications will be 48, which is 3/16 of the original 256.

This strategy for reducing the number of computations can be traced back long before Cooley and Tukey's work to the astronomer Carl Friedrich Gauss. Gauss wanted to calculate asteroidal and cometary orbits from only a few observations. After discovering a solution, he found a way to reduce the complexity of the calculations based on principles similar to those of the fast Fourier transform. In an 1805 paper describing the work, Gauss wrote: "Experience will teach the user that this method will greatly lessen the tedium of mechanical calculation." Thus, the challenge of celestial motions not only gave us calculus and the three laws of motion but also stimulated the discovery of a modern computing tool.

Physicists and engineers, indoctrinated with complex algebra early in their education, have become comfortable with the representation of sinusoids. The convenience of representing the Fourier transform as a complex function lets us forget that the underlying sinusoidal components are real and not necessarily complex. This habit of thought has obscured the significance of and retarded the adoption of a transform similar to Fourier's that was conceived by Ralph V. L. Hartley in 1942.

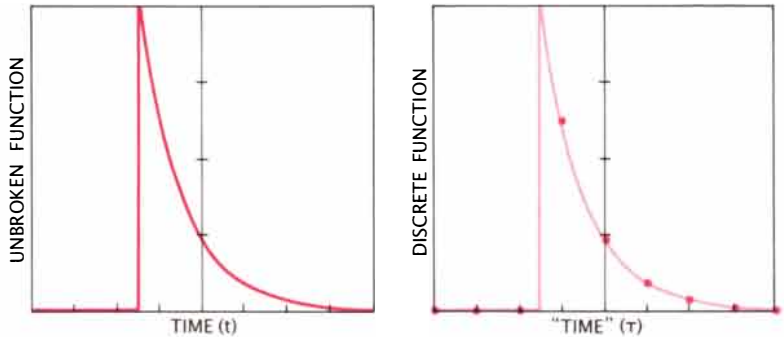
Working in the research laboratory of the Western Electric Company, Hartley directed the early development of radio receivers for a transatlantic radiotelephone and invented the Hartley oscillating circuit. During World War I Hartley investigated how a listener, through mechanisms in the ear and brain, perceives the direction from which a sound emanates. Working at Bell Laboratories after the war, Hartley was the first to formulate an important principle of information technology that states that the total amount of information a system can transmit is proportional to the product of the frequency range the system transmits and the time during which the system is available for transmission. In 1929 Hartley gave up the direction of his group because of illness. As his health improved he devoted himself to the theoretical studies that led to the Hartley transform.

The Hartley transform is an alternative means of analyzing a given function in terms of sinusoids. It differs from the Fourier transform in a rather simple manner. Whereas the Fourier transform involves real and imaginary numbers and a complex sum of sinusoidal functions, the Hartley transform involves only real numbers and a real sum of sinusoidal functions.

In 1984 I developed an algorithm

THE FOURIER AND HARTLEY TRANSFORMS

The Fourier and Hartley transforms convert functions of time into functions of frequency that encode phase and amplitude information. The graphs below represent the unbroken function $g(t)$ and the discrete function $g(\tau)$, where t is time and τ is a number designated at each data point.



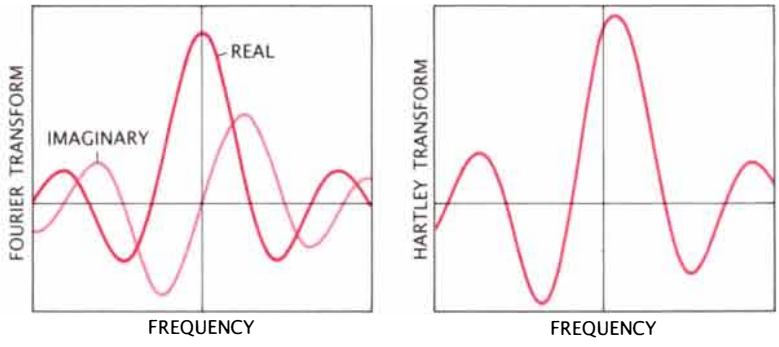
Both functions start at zero, jump to a positive value and then decay exponentially. The definition of the Fourier transform for the unbroken function is an infinite integral, $F(f)$, whereas the definition for the discrete function is a finite sum, $F(v)$.

$$F(f) = \int_{-\infty}^{\infty} g(t)\ (\cos\ 2\pi ft - i \sin\ 2\pi ft)\ dt \qquad F(v) = \frac{1}{n} \sum_{\tau=0}^{n-1} g(t)\ (\cos\ 2\pi v\tau - i \sin\ 2\pi v\tau)$$

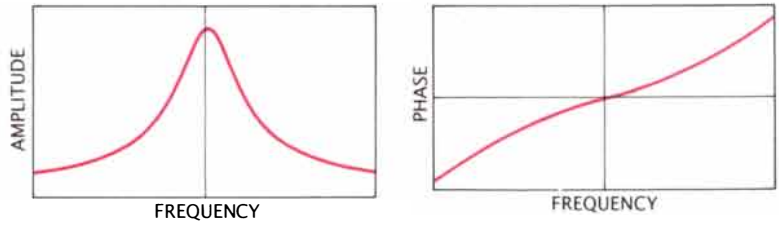
Here f is frequency, v is related to frequency, n is the total number of samples and i is the imaginary number equal to the square root of -1. The integral representation is more suited to theoretical manipulations, whereas the finite-sum representation is more suited to computer applications. The Hartley transform and discrete Hartley transform have similar definitions.

$$H(f) = \int_{-\infty}^{\infty} g(t)\ (\cos\ 2\pi ft + \sin\ 2\pi ft)\ dt \qquad H(v) = \frac{1}{n} \sum_{\tau=0}^{n-1} g(t)\ (\cos\ 2\pi v\tau + \sin\ 2\pi v\tau)$$

Even though the only notational difference between the Fourier and Hartley definitions is a factor $-i$ in front of the sine function, the fact that the Fourier transform has real and imaginary parts makes the representations of the Fourier and Hartley transforms quite different. The discrete Fourier and discrete Hartley transforms have essentially the same shape as their unbroken counterparts.



Although the graphs look different, the phase and amplitude information that can be deduced from the Fourier and Hartley transforms is the same, as shown below.



Fourier amplitude is the square root of the sum of the squares of the real and imaginary parts. Hartley amplitude is the square root of the sum of the squares of $H(-v)$ and $H(v)$. Fourier phase is the arc tangent of the imaginary part divided by the real part, and Hartley phase is 45 degrees added to the arc tangent of $H(-v)$ divided by $H(v)$.

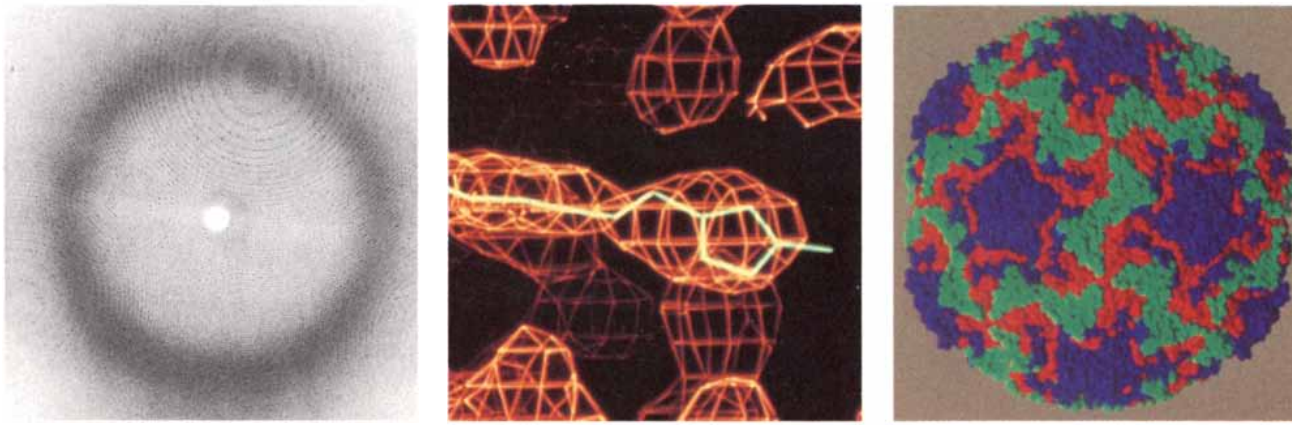
for a fast Hartley transform. The difference in computation time between the fast Hartley transform and the fast Fourier transform depends on the computer and the programming language and style. If these factors are kept constant and no oversights are made in the programming, programs for the fast Hartley transform run more quickly than those for the fast Fourier transform. Although both programs need the same amount of time to retrieve the data, provide for trigonometric functions and perform other preliminaries, the time spent on the stages of the Hartley transform is half that required by the Fourier.

It was not clear at first, however, that the Hartley transform provided the same information as the Fourier transform. Therefore, when the first programs were developed for computing the Hartley transform, an extra step was provided to convert it to the more familiar Fourier. Workers soon realized, however, that intensities and phases can be deduced directly from the Hartley transform without the need for the additional step. Further reflection revealed that either kind of transform furnishes at each frequency a pair of numbers that represents a physical oscillation in amplitude and phase.

Yet another reservation about the Hartley transform was that the Fourier transform described physical phenomena more naturally. Many phenomena, such as the response of a simple system to vibration, are commonly described by a complex sum of sinusoidal functions, which is the hallmark of the Fourier transform. It might seem, therefore, that Fourier transforms are more suitable for describing the behavior of nature.

Such a conclusion is in fact more a reflection of our mathematical upbringing than it is of nature. After all, when physical objects are measured, they provide data in real numbers, not complex ones.

The advent of the fast Hartley transform has made obsolete certain adaptations of the fast Fourier transform, such as those used for eliminating noise from digitally recorded music. These adaptations require two programs: one of them transforms real functions into the complex Fourier domain, whereas the other converts complex functions from the Fourier domain into real functions. High-frequency noise in digitally recorded music can be eliminated by filtering out portions of the transform produced by the first program. The second program then converts the changed trans-



FOURIER ANALYSIS can transform X-ray diffraction patterns into molecular models. X rays scatter off the electrons in a virus, for example, to produce patterns on film (*left*). These patterns represent part of the Fourier transform of the virus's molecular structure. If the process of transformation is reversed, the distribution of electrons, and therefore atoms, can be deduced (*middle*). From these distributions, models of the virus are made (*right*). Here colors indicate different proteins.

form back into an improved musical signal. Although these ingenious programs run individually at speeds rivaling the fast Hartley transform, a single Hartley program suffices for both transforming a real function into a Hartley transform and converting the transform, after the desired filtering, back to a real function. Therefore, extra computer memory for storing two programs is not required.

In the most general terms, Fourier and Hartley transforms have been applied in fields that contend with fluctuating phenomena. Their field of application is thus very broad indeed.

Many applications exist in biology. In fact, the double-helix form of DNA was discovered in 1962 through X-ray diffraction techniques and Fourier analysis. A beam of X rays was focused on a crystal of DNA strands, and the X rays were diffracted by the molecules of the DNA and recorded on film. This diffraction pattern provided the amplitude information of the crystal structure's Fourier transform. The phase information, which the photographs alone did not provide, was deduced by comparing the DNA diffraction pattern with patterns produced by similar chemicals. From the X-ray intensity and phase information in the Fourier transform, biologists worked back to a crystal structure—the original function. In recent years, X-ray diffraction studies combined with such "reverse" Fourier analysis have revealed the structure of many other biological molecules and more complex structures, such as viruses.

The National Aeronautics and Space Administration improves the clarity and detail of pictures of celestial objects taken in space by means of Fourier analysis. Planetary probes and earth-orbiting satellites transmit images to the earth as a series of radio impulses. A computer transforms these impulses by Fourier techniques. The computer then adjusts various components of each transform to enhance certain features and remove others—much as noise can be removed from the Fourier transforms of recorded music. Finally, the altered data are converted back to reconstruct the image. This process can sharpen focus, filter out background fog and change contrast.

The Fourier transform is also valuable in plasma physics, semiconductor physics, microwave acoustics, seismography, oceanography, radar mapping and medical imaging. Among the many applications in chemistry is the use of the Fourier-transform spectrometer for chemical analysis.

Fourier analysis has proved valuable in my own work in two-dimensional imaging. In 1956 I stumbled on a "projection slice" theorem that yielded a way to reconstruct images from strip integrals, a problem now widely known as tomographic reconstruction. Later, I hit on the "modified backprojection" algorithm, now universally used in computer-assisted X-ray tomography, or CAT scanning.

I was also interested in reconstructing images based on data from radio astronomy. I wanted to pinpoint sources of radio waves on the sun's surface, so I applied transform methods to the design of a scanning radio telescope that made daily microwave temperature maps of the sun for 11 years. The methods led to the first antenna with a beam sharper than the resolution of the human eye and have since diffused into general antenna technology. NASA commended the maps of the sun for contributing to the safety of the lunar astronauts.

I have also applied the Hartley transform to other studies. Recently my colleague John D. Villasenor and I described an optical method for finding the Hartley transform, a development that enables Fourier phase and amplitude to be encoded in a single real image. We have also developed a device that constructs the Hartley transform using microwaves. I am now writing papers on solar physics in which transform techniques underlie new ways of analyzing data from sunspot counts and from the thickness of sedimentary layers on the earth.

The wide use of Fourier's method and related analytical techniques makes what Lord Kelvin said in 1867 just as true today: "Fourier's theorem is not only one of the most beautiful results of modern analysis, but it may be said to furnish an indispensable instrument in the treatment of nearly every recondite question in modern physics."

FURTHER READING

JOSEPH FOURIER: THE MAN AND THE PHYSICIST. John Herviel. Clarendon Press, 1975.

THE FOURIER TRANSFORM AND ITS APPLICATIONS, SECOND EDITION, REVISED. Ronald N. Bracewell. McGraw-Hill Book Company, 1986.

THE HARTLEY TRANSFORM. Ronald N. Bracewell. Oxford University Press, 1986.

OPTICAL PHASE OBTAINED BY ANALOGUE HARTLEY TRANSFORMATION. John Villasenor and R. N. Bracewell in *Nature,* Vol. 330, No. 6150, pages 735–737; December 24, 1987.

The Channeling of Electrons and Positrons

Charged particles traveling along a crystal's planes of symmetry behave strangely: they interact with sheets or strings of nuclei rather than with single atoms

by Allan H. Sørensen and Erik Uggerhøj

When charged particles hit a solid object, they generally scatter more or less randomly as they collide with individual atoms. When the target is a crystal, however, and a narrow, focused beam of particles strikes it from a particular angle, individual scattering effects give way to collective ones, in which the particles are guided by long strings or by planes of atoms. Such channeling may enable a particle to penetrate much farther into the crystal than it would otherwise; the process also endows the radiation the particle emits with specific characteristics and dramatically enhances or suppresses reactions between the particle and nuclei in the target crystal.

Channeling offers a way to probe crystal structure and a new arena in which to study the interactions of radiation with matter. The application of channeling principles may result in new sources of X rays and gamma rays consisting of electron or positron accelerators trained on carefully aligned crystals of germanium, silicon or other elements. Similar crystals set up to absorb high-energy radiation could act as gamma-ray detectors, capable of far better angular resolution than those existing today.

To a charged particle impinging on a crystal along one of the crystal's axes of symmetry, the atoms in the crystal appear to be lined up much like pearls on a string. Collisions with atoms in any given string are strongly correlated: if the particle passes close to one atom, it must also pass close to neighboring atoms in that same string. As a result the string acts as a whole, deflecting the charged particle as if the discrete charges of the atoms in the string had been smeared out uniformly along its length. Similarly planes of atoms in a crystal can channel particles by acting as flat sheets between which the particles are confined.

Positively charged particles that are channeled will travel about four to five times farther into a crystal before dissipating their energy than they would otherwise. In addition, certain kinds of interactions, such as close encounters with nuclei of the target crystal, will be selectively suppressed. Mutual repulsion confines a positively charged particle mostly to the "open" spaces between strings or planes, thereby making reactions less likely. For negatively charged particles, the picture is almost precisely the opposite: they travel primarily along the lines of positively charged nuclei, and so the probability of close-encounter reactions can be boosted as much as sixfold.

Although the effect of channeling on penetration depth and projectile-target interactions is quite strong, its effects on the radiation emitted by projectiles are perhaps even more distinctive. Particles traversing an amorphous solid produce bremsstrahlung radiation—broadband emission of photons at all energies, up to the energy of the particle itself—as they decelerate. But channeled particles emit intense beams of photons at specific energies. In some cases sharp emission lines appear, much like those of excited electrons in atoms.

Because lighter particles emit far more radiation than heavier ones, most of the current interest in channeling focuses on positrons and electrons rather than on particles such as protons, alpha particles or positive and negative pions. In classical nonrelativistic terms, the energy emitted per unit of time by a charged particle is proportional to the square of its acceleration, which is the applied force divided by the mass. The electrostatic forces affecting the proton and electron are of equal magnitude, but because of the difference in mass the energy emitted by a channeled proton is six orders of magnitude less than that emitted by an electron. Certain relativistic effects further magnify the difference at any given energy.

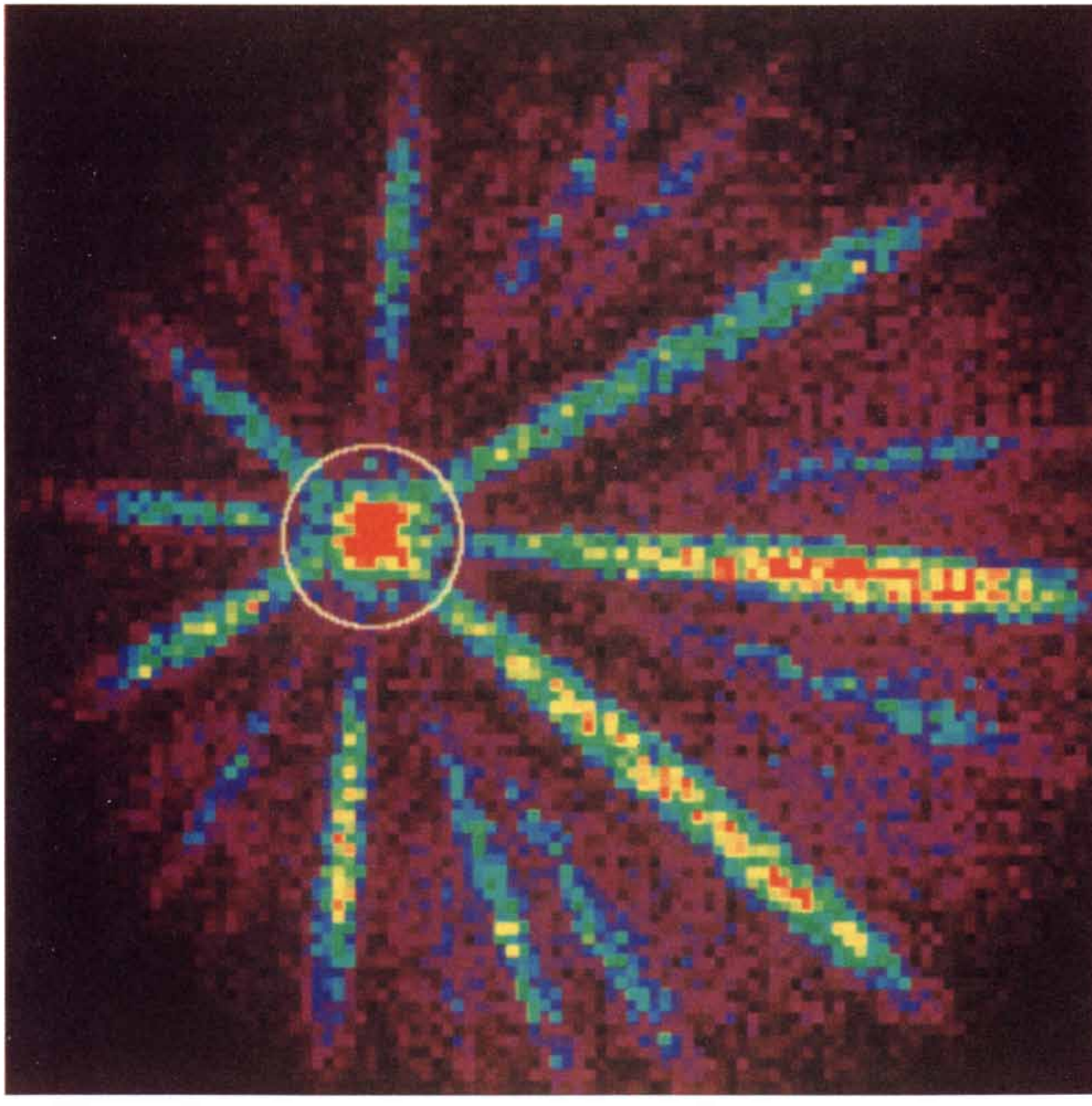
In the classical metaphor for planar channeling, the positively charged particle is a ball rolling down a long gutter that represents the space between two adjacent planes of atoms. There is essentially no resistance to the motion of the ball along the gutter; the gutter affects the ball's motion only in the transverse direction. A positively charged particle will be repelled by the positively charged nuclei of the planes that define the walls of the gutter. The energy of the ball can therefore be divided into two terms: a large one corresponding to

ALLAN H. SØRENSEN and ERIK UGGERHØJ are on the faculty of the Institute of Physics at the University of Aarhus in Denmark, where they study channeling radiation and other phenomena associated with the penetration of charged particles through matter. Sørensen began work on channeling while he was a graduate student during the early 1980's; after finishing his doctorate, he began collaborating with Uggerhøj's experimental team at CERN (the European laboratory for particle physics near Geneva). Uggerhøj has spent much of his career investigating channeling; he carried out some of the first channeling experiments in the mid-1960's and pioneered the study of electron and positron channeling. Both Uggerhøj and Sørensen have recently studied the production of particle-antiparticle pairs by gamma rays striking a crystal near its axes, an effect that could be harnessed for high-resolution orbiting gamma-ray telescopes.

the kinetic energy of the motion along the gutter and a smaller one corresponding to the sum of the transverse kinetic energy and the potential energy caused by the ball's interaction with the gutter. As the ball rolls from side to side, it exchanges transverse kinetic energy and potential energy.

The closest approach between the projectile and a nucleus—the point on the gutter wall where the ball turns and starts rolling downward—depends on the transverse energy. For a reaction to occur the turning point must be very close to the nucleus or, in the rolling-ball analogy, very near the top of the gutter. The yield of a process that requires close contact between a projectile and an atomic nucleus (such as a nuclear reaction) will lessen as the maximum angle between the projectile path and the crystal plane decreases; below a critical angle the reaction will be suppressed almost completely.

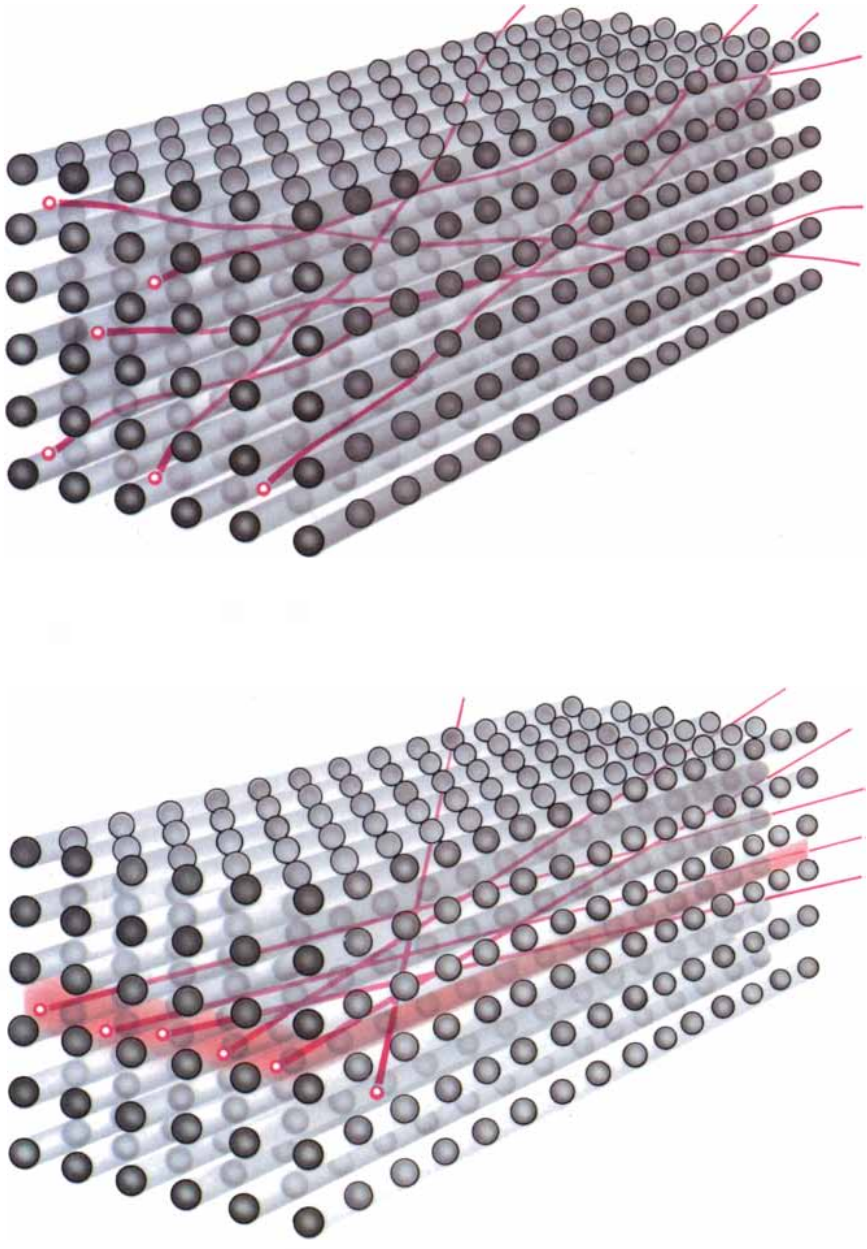
From the height of the gutter—the transverse energy required for a projectile to cross from one planar channel in a crystal to another—and the energy of the incoming particle, it is simple to calculate the critical angle below which the particle will remain in a single channel. For protons impinging on a tungsten crystal at an energy of a million electron volts (MeV), for example, the critical angle is about .4 degree; for the same protons on silicon it is about .2 degree. At smaller angles of incidence essentially all the protons are channeled; at larger angles they behave much as if they were passing through an amorphous solid.



CHANNELED PROTONS stream through a thin germanium crystal in a view along an axis of symmetry. The colors indicate how many protons survive passage without a significant change of direction: protons channeled along an axis are scattered least (*red*); those channeled by crystalline planes are scattered somewhat more. Travel across lines or planes of atoms tends to cause the strongest scattering. Most protons that pass through the crystal in such directions change their paths (*blue, purple*).

In axial channeling the description in terms of gutters is no longer applicable. The particle travels along a crystalline axis, but transverse motion proceeds in two directions perpendicular to the direction of travel instead of one. A positively charged particle, for example, can be pictured as rolling on a surface with sharp peaks representing the lines of atoms and broad plains between them. The ball is not confined to the space between adjacent peaks; instead, it can wander anywhere on the plain. Because the interaction between a projectile and a single line of atoms is generally stronger than the interaction between a projectile and a plane, the critical angle for axial channeling in a given crystal is about three times larger than that for planar channeling: 1.6 degrees in tungsten; .6 degree in silicon.

For negatively charged particles the picture of channeling is inverted. The force between the atomic nuclei and the incoming projectiles is attractive rather than repulsive. The gutters of planar channeling have been turned upside down, yielding sharp troughs corresponding to the planes of atoms instead of broader troughs representing the spaces between them. The peaks of axial channeling are now deep, circular valleys in which a particle may become trapped. Because a particle with low transverse energy is forced to remain in the vicinity of target nuclei, reactions that depend on close encounters are enhanced rather than suppressed.



LINES AND PLANES OF ATOMS in a crystal channel incoming particles. In axial channeling (*top*) particles impinging at an angle near a crystalline axis are guided by glancing collisions with lines of atoms, so that they travel only between the lines. In planar channeling (*bottom*) particles interact with sheets of atoms; they are free to move across a plane but not from one plane to another. The illustration shows the channeling of positively charged particles, which are repelled by the nuclei in the crystal. Negatively charged particles are channeled near nuclei.

Although the ball-and-gutter model for channeling gives a good intuitive understanding of how charged particles interact with the atomic planes of crystals, it is not actually correct. The microscopic world of charged particles and atomic nuclei is governed by quantum mechanics rather than classical physics. In general, charged particles are not well-localized objects that slalom back and forth between planes of atoms; instead, they occupy quantum states corresponding to specific values of longitudinal and transverse energy. The probability of finding a particle of a given transverse energy in a particular section of transverse space is generally quite different in a quantum-mechanical and a classical model. How then does the classical picture describe channeling so well to a first approximation? What corrections must be made to the details of the classical picture to account for the underlying quantum-mechanical description?

For heavy particles such as protons or alpha particles, classical physics describes the channeling quite well. The physical separation between quantum states depends inversely on mass, so that the same space can encompass more quantum states of a heavy particle than of a light one. For example, an atom consisting of a proton and an antiproton instead of a proton and an electron would have a thousandth the radius of a standard hydrogen atom. More than 30 energy levels of the proton-antiproton system would fit within the radius of the lowest electron energy level. Systems that have many closely spaced quantum levels behave much like classical ones: for the higher energy levels the classical and quantum probability distributions (the chances of finding a particle in a given location) are similar, and the transition between closely spaced quantum states cannot be distinguished from continuous behavior. This is Niels Bohr's well-known correspondence principle.

The same arguments that hold for atoms also hold for projectiles channeling through crystals: massive projectiles behave according to the laws of classical mechanics. Their mass enables them to pack more quantum states into the transverse space. Moreover, because these massive particles are for the most part positively charged, they travel in the spaces between lines or planes of nuclei, whereas negatively charged particles travel in the much smaller region very close to the nuclei. The extra space makes possible even more quantum states.

In contrast, an electron whose energy is a few MeV undergoing axial channeling in a silicon crystal may have a choice between only 10 or fewer quantum states representing its transverse motion. For planar channeling, where the gutter is much shallower, the electron may be restricted to a single state of transverse motion. The limited number of quantum states has a number of counterintuitive consequences: for example, for low-energy states the probability density—the chance of finding the electron at any given position within the channel—differs markedly from classical predictions. The quantum structure is also apparent in the spectrums of radiation emitted by channeled electrons. Channeled electrons emit radiation at specific wavelengths that correspond to transitions from one transverse energy state to another.

Even when the electron is limited to a few quantum states, however, some of its behavior can still be explained classically. For example, classical and quantum predictions match closely for the scattering of electrons by close encounters with nuclei. In these scattering events the electrons are deflected from their original paths by angles much larger than the critical channeling angle. The classical and quantum predictions match because the total probability density of electrons in regions close to nuclei does not differ much from the classical probability. Thus, even when quantum effects clearly dominate the phenomena being studied, classical mechanics can sometimes be used to model them.

The same basic principles that govern channeling at moderate energies also hold at higher projectile energies, where the effects of special relativity become important. Relativistic effects come into play when a particle moves at speed near that of light. At such speeds the kinetic energy of a projectile becomes comparable to or larger than its rest energy (the energy equivalent of its rest mass). For protons this threshold is about one billion electron volts (GeV); for electrons it is approximately 500,000 electron volts.

For a particle moving through a crystal at relativistic speed, the number of quantum states available depends not only on its rest mass and charge but also on the Lorentz factor, gamma, which measures the ratio of the total energy of a particle to its rest energy. The number of available states increases with gamma because of the foreshortening caused by relativistic travel. Objects traveling close to the speed of light appear shorter in the direction of motion by a factor of gamma (this is known as the Lorentz-Fitzgerald contraction). As the channeled particle "sees" things, the crystal is rushing by at nearly the speed of light. The crystal appears contracted, and so the apparent spacing between atoms is smaller. This contraction increases the amount of charge per unit length in the strings of atoms channeling the particle, increasing the depth of the potential-energy well and thereby making room for additional transverse energy states. This relativistic increase in the number of quantum states can also be attributed to the relativistic increase (in the laboratory frame of reference) of the particle's mass. At high energies a light particle behaves as if it were considerably heavier, with a resulting increase in the number of quantum states. Paradoxically, then, electrons with very high energy (GeV or more) also behave classically, because they have many quantum states from which to choose.

Relativity affects not only the spacing of quantum states for channeled particles but also the radiation emitted as particles make transitions from one state to another. If the transverse energy were only involved, the photons emitted by an electron with energies of a few MeV would typically be in the visible or the infrared region of the spectrum. As the electron energy increased, increasing the available number of quantum levels and reducing their spacing, the energy of the emitted photons would tend to decrease. Relativistic effects, however, shift the wavelength of the photons emitted by such electrons into the X-ray portion of the spectrum.

The relativistic underpinnings of this shift can be understood by looking at the radiation process from the particle's point of view. In the particle's reference frame there is no longitudinal motion, only transverse motion. The transverse potential, however, is magnified by a factor of gamma because of the relativistic contraction that reduces the spacing between atoms. Therefore, a transition in transverse energy is magnified by the same factor as compared with the laboratory frame. Furthermore, moving from the particle's reference frame back to the laboratory frame results in another energy increase: because the particle is moving, the radiation it emits is Doppler-shifted. Viewed from the front, the photons emitted by the particle are shorter in wavelength and therefore higher in energy by a factor of as much as two times gamma. Hence, the total magnification of the energy is twice gamma squared for photons emitted straight ahead—a factor of about 200 for a 5-MeV electron. This is quite sufficient to turn an infrared photon into an X-ray photon.

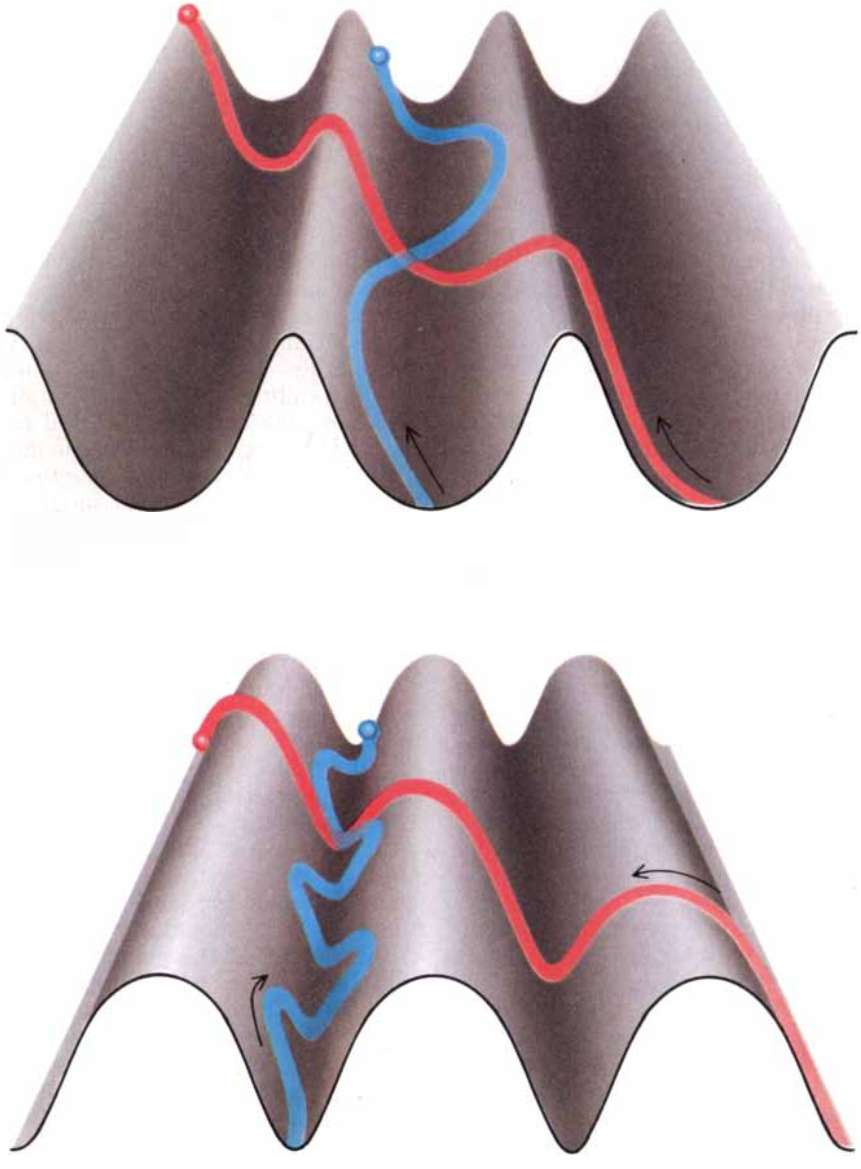
In addition to the relativistic energy shift, radiation from a channeled particle experiences what is called relativistic beaming. Although in the rest frame of the channeled electron photons are emitted about equally in all directions, the emission observed in the laboratory is strongly biased in the forward direction. Half of the photons are typically emitted within an angle of 1/gamma radians (2π radians correspond to the 360 degrees of a circle) relative to the channeled particle's direction of travel. Because they are Doppler-shifted, these photons carry most of the radiated energy.

The radiation emitted by electrons with energies of a few MeV channeled in crystals of light elements such as carbon or silicon gives rise to a series of discrete spectral lines, typically at X-ray wavelengths. When the electron energy is increased, the line structure of the emissions gradually disappears. The number of transverse quantum states increases because of the increased depth of the potential-energy well in which the electrons travel, making the number of possible transitions increase as well. Because the many lines all have a finite width, they tend to blend together into a continuum. This change in the spectrum mimics the change from quantum-mechanical to classical behavior in the energy levels themselves.

Radiation emitted from a process at still higher energies—GeV-energy positrons undergoing planar channeling—has attracted special attention in the physics community, because the model describing its production is very simple and the radia-

tion itself is almost perfectly monochromatic. The force between a set of crystal planes and the positron depends linearly on the displacement of the positron from the center of the channel. The restoring force leads to a regular transverse oscillation whose frequency is almost entirely independent of the transverse energy. The positrons act like tiny pendulums: their period does not depend on the length of their excursions from the center line of their travel. As a result all the photons emitted in the forward direction have almost precisely the same energy: somewhere between 10 and 100 MeV for positrons between 1 and 10 GeV in a silicon crystal, depending on the exact energies of the positron and the channeling plane.

The intensity of the channeling radiation is greater than that of conventional bremsstrahlung radiation. The enhancement results from the coherent scattering that takes place at small angles. Because the channeled particles are traveling along the lines of atoms, one glancing collision—and the photon emission it triggers—will be followed by many more. In contrast, particles traveling through an amorphous medium undergo collisions and scattering at many different angles and produce broad-spectrum, less intense emission. Channeling radiation is roughly 50 times more intense than conventional bremsstrahlung radiation at gamma-ray energies of up to a few GeV. It is important to remember, however, that this enhancement pertains only to photons with energies of up to a few percent of the impact energy; bremsstrahlung photons may be emitted at any energy up to the kinetic energy of the projectile. Channeled particles emit less radiation at very high energies than do unchanneled particles.



BALL-AND-GUTTER ANALOGY describes the behavior of particles channeled by crystalline planes. Balls experience no resistance along the direction of travel, but their transverse motion is confined. A channeled particle (*blue*) travels up the side of the gutter, exchanging kinetic energy for potential energy; eventually its transverse motion reverses. Particles with too much transverse energy to be channeled (*red*) cross from one gutter to the next. For positively charged particles (*top*) the gutters have flat bottoms and steep sides; for negatively charged ones the opposite is true (*bottom*). Nuclei are spaced along the walls of the gutters for positive particles and in the troughs for negative ones. Channeling suppresses reactions between positive particles and nuclei, but it enhances them for negative particles.

When the energy of incoming electrons or positrons rises still higher, to levels above a few GeV, the classical description of channeling radiation becomes invalid again. At these energies there are easily enough transverse quantum states to approximate a continuum, but a different assumption breaks down instead. The classical treatment of the radiation process relies on the particle path's remaining essentially unaffected as the particle emits photons. Because of the relativistic effects discussed earlier, however, the average energy of emitted photons increases as the square of the Lorentz factor. At high energies a single radiated photon can carry away an appreciable fraction of the projectile's energy, thereby altering its path significantly.

On the other hand, at these very high energies another effect comes into play that simplifies the radiation calculations. At low energies the cone within which photons are emitted measures a few degrees, much wider than the fraction of a degree through which the path of the electron or positron varies as it travels through the channel between lines or planes of atoms. At high energies, in contrast, relativistic beaming narrows the emission cone, so that it is much narrower than the angular excursion of the projectile. Instead of sweeping a broad beam through a narrow angle, so that radiation from different segments of the path overlaps and cannot be distinguished, the high-energy projectile radiates a pinpoint beam that sweeps through a broader angle. The contributions of any given point on its path are independent.

As a result the projectile path can be

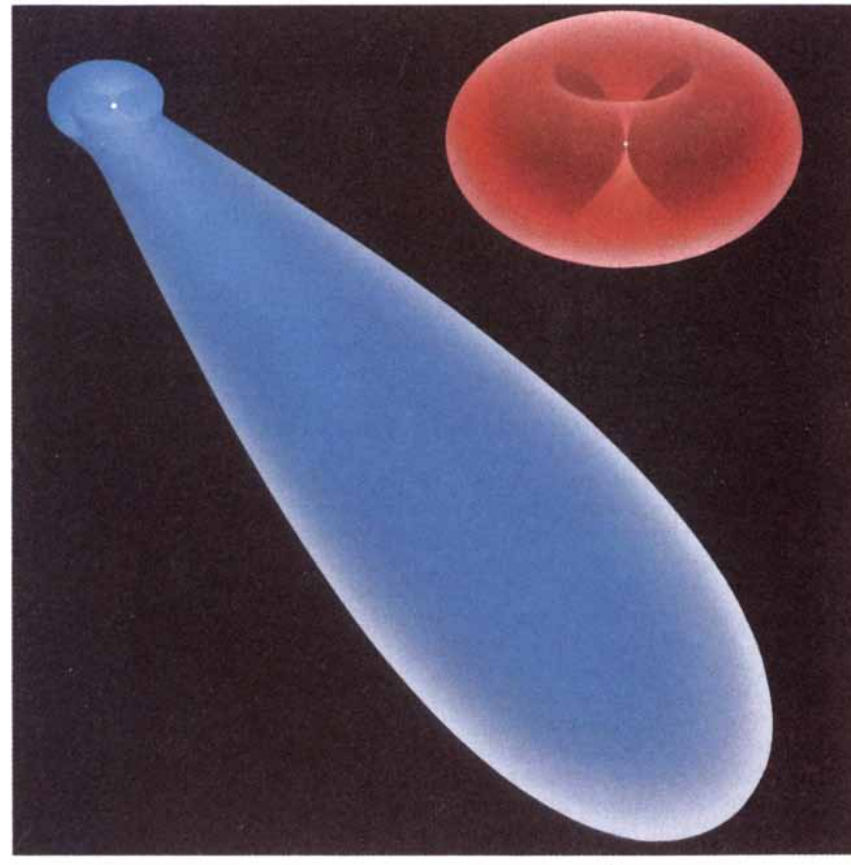
considered as a series of small segments, with the radiation for each segment being calculated independently. Furthermore, within each segment of the path the electromagnetic field of the atoms in the crystal is essentially constant, and the problem of photon emission by a charged particle moving in a strong, constant electromagnetic field is very well understood. This constant-field approach is applicable at very high energies for unchanneled particles as well. One interesting prediction of the model is a substantial enhancement of radiation above the levels observed for amorphous targets whenever a beam of particles impinges along a direction within .01 to .1 degree of a crystal axis, depending on the material of the crystal.

The ability of a high-energy channeled particle to radiate almost all of its kinetic energy in a single photon gives rise to an effect that does not take place at lower energy: the enhanced creation of particle-antiparticle pairs by photons. In the presence of an external electromagnetic field such as the electric field inside a crystal, an energetic photon can convert into an electron-positron pair. About 1 MeV goes into creating the mass of the pair, and the rest furnishes the electron and positron with kinetic energy. Pair creation occurs in amorphous targets as well, but the effect is enhanced when photons move along a crystalline axis because the lines or planes of atoms present a stronger, more coherent electromagnetic field. The enhanced pair production becomes apparent above an energy threshold that depends on the channeling crystal; along the most efficient axis in a germanium crystal, for example, this threshold is about 40 GeV.

Because the photon is electrically neutral, there are no focusing effects along the crystalline axes; the photon is not channeled in the sense of having its path directed along a crystalline axis. As a result there is no angular threshold like the critical angle for channeling seen for electrons and positrons. Instead, enhanced pair production by photons aligned with an axis shows the same angular distribution as the enhancement over bremsstrahlung radiation found for high-energy charged particles aligned with crystalline axes. In germanium, pair production is readily distinguishable at angles less than about .1 degree from a crystalline axis.

Channeling radiation has a number of applications as a research tool for studying solid-state structure and other phenomena. For MeV electrons and positrons the discrete emission lines constitute a fingerprint of the different states of projectile motion. These states in turn provide information about the internal structure of the crystal in which the channeling takes place. In such "channeling spectroscopy" the positions of photon peaks are sensitive to specific details in the distribution of target electrons among the atoms in the crystal lattice. In areas with high electron density the atomic nuclei will be shielded, and so the potential-energy well will be shallower, and quantum states that tend to place electrons in those regions will be shifted upward in energy; in areas with low density the potential will be deeper. In addition, the most likely transverse position of the channeled particle will vary depending on its quantum state. Consequently experiments can be performed to scan regions at specific distances from nuclei by controlling the particle's transverse energy.

The positions of the peaks can yield information on the magnitude of the thermal vibrations of the target atoms, and analysis of the width of each line can reveal the extent to which those vibrations are either erratic or correlated. All emission lines have a finite width because of the uncertainty principle, which implies that electron energy levels cannot be known precisely, because electrons spend only a limited time in any given state. The uncertainty in energy levels translates to uncertainty in the difference between two levels and thus in the energy of the photon emitted in the transition. Electrons at energies of about 5 MeV take about three femtoseconds to pass through a micron-thick crystal in the laboratory frame of reference and even less (by a factor of gamma) in their own reference frame. After accounting for the Doppler shift, which magnifies both total energy and line width, the uncertainty is a few hundred electron volts. But if a channeled electron is scattered by an atom whose erratic thermal motion places the atom directly in its path, as is often the case, the lifetime of its channeling quantum state will be much



RELATIVISTIC BEAMING transforms the pattern of radiation emitted by a channeled particle. In the particle's reference frame, photons are emitted in all directions (except straight up and down), but in the laboratory frame the radiation is seen primarily in a narrow cone along the particle's direction of motion. The Doppler shift increases the observed energy of the photons emitted within the cone.

shorter than the flight time through the crystal, and the corresponding line width will be greater. In most cases it has been found that the vibrations of one atom are correlated with those of its neighbors, so a collision with one means a collision with others, further shortening lifetimes and broadening lines. The exact degree of correlation can be deduced from the degree to which the lines are broadened. It is not yet clear, however, whether channeling radiation will be competitive with other, better established techniques for studying crystal structures.

Perhaps the most promising prospective application of channeling radiation is as an intense, tunable source of X-ray and gamma radiation. X rays can be generated using the quantum peaks that appear at electron energies of a few MeV, and the emissions from planar-channeled positrons at a few GeV can serve for generating gamma rays. Because planar-channeled particles vibrate only perpendicular to the channeling planes, the resulting radiation will be highly polarized, thus making it useful for investigations of polarization-dependent processes and for experiments where a signal must be extracted from an unpolarized background.

Channeling radiation is now beginning to find use as an X-ray source. The photon yield is fairly low, however: typically one photon per thousand incident electrons, which makes it difficult to produce very intense beams. As a gamma-ray source, on the other hand, channeling appears more promising because it yields roughly one photon per incident positron, and there are few competing sources.

Some groups have proposed using gamma rays generated by channeling to study the energy levels of fissionable heavy nuclei. Gamma rays of known energy can stimulate a nucleus into excited states in which fission becomes more likely; the tunable radiation from channeled positrons is well suited to this purpose. The same radiation can also serve to stimulate light elements such as deuterium or beryllium to emit neutrons with high efficiency. Because the channeling radiation is highly collimated, the effective radius of the neutron source is the same as the radius of the particle beam producing the radiation—as small as a few microns. The resolution is high in time as well as space: a pulsed particle beam can produce neutron pulses as short as a few nanoseconds. The resulting neutron source can be used to study many time-dependent phenomena.

Many other applications for channeling radiation have been proposed. One very interesting possibility is a channeling laser. Lasers work by creating population inversions, in which higher-energy states contain more electrons than lower-energy states. A single electron dropping from a high state to a low one emits a photon that stimulates a cascade of transitions and emissions from the other electrons. In a channeling laser, electrons in states of high transverse energy would outnumber those in states of lower energy; the resulting laser emission would take place at X-ray wavelengths. Unfortunately a channeling laser would require a very high electron current, on the order of 10,000 amperes per square centimeter, which would rapidly heat the crystal and damage it.

High-energy directional processes (above about 40 GeV) offer their own unique applications. Some of these might be developed in the near future, such as direction-sensitive detectors for high-energy photons. Gamma-ray telescopes based on enhanced pair-creation effects could be flown aboard satellites. A gamma-ray photon with an energy above 100 GeV has a high probability of producing an electron-positron pair in a germanium crystal between two and three millimeters thick, provided the photon enters within about .05 degree of a major axis. The electron and positron will then interact with the crystal and radiate high-energy photons that in turn will create more pairs, resulting in many lower-energy electrons, positrons and photons. At high impact energy along the most effective axis of a germanium crystal, such cascades develop in approximately one centimeter, versus up to 20 centimeters for an unaligned impact. Replacing germanium with denser elements of higher atomic number, such as tungsten, would extend the useful range of this effect to lower photon energies. Such gamma-ray detectors could be exploited to improve the angular resolution of gamma-ray maps of the sky by orders of magnitude. Current gamma-ray telescopes can only determine the position of a source to within one degree. Some materials, such as silicon, can serve not only to stimulate pair creation but also to detect electron-positron pairs; a telescope based on such crystals would be very simple to construct.

Some questions remain about channeling and related processes at GeV energies. Although measured pair-creation rates agree very well with theoretical predictions of variation with energy and angle, theoreticians have had trouble predicting the spectrum of radiation emitted by high-energy channeled particles. In one experiment done with 150-GeV electrons and positrons channeled through a thin germanium crystal, for example, electrons produced more photons than positrons did—which was expected because the electrons travel closer to nuclei, where the electric fields are stronger—but most of the additional photons were gathered in a strong peak at 85 percent of the electron energy. This peak was completely unexpected.

Theoreticians have now deduced the mechanism responsible: electrons emit multiple photons that are recorded as a single photon with the combined energy of the component photons. This "radiative cooling" process reduces the transverse energy of the electrons, so that they spend even more time in the vicinity of nuclei, further intensifying the emission. For now, it appears that no new principles need to be invoked to explain this phenomenon.

To date, experimental data have been rather scarce, and so the theory has not been fully tested. Other equally unexpected phenomena may appear as electrons and positrons are channeled at even higher energies. Further experiments are urgently needed to help determine exactly what models govern the behavior of channeled particles in the highest-energy regimes.

FURTHER READING

CHANNELING AND RELATED EFFECTS IN THE MOTION OF CHARGED PARTICLES THROUGH CRYSTALS. Donald S. Gemmell in *Reviews of Modern Physics,* Vol. 46, No. 1, pages 129–227; January, 1974.

CHANNELING RADIATION. J. U. Andersen, E. Bonderup and R. H. Pantell in *Annual Review of Nuclear and Particle Science,* Vol. 33, pages 453–504; 1983.

CHANNELLING AND CHANNELLING RADIATION. Allan H. Sørensen and Erik Uggerhøj in *Nature,* Vol. 324, No. 6102, pages 311–318; January 22, 1987.

PARTICLE PRODUCTION IN CRYSTALS. Allan H. Sørensen and Erik Uggerhøj in *Nature,* Vol. 327, No. 6123, pages 558–559; June 18, 1987.

RELATIVISTIC CHANNELING. Edited by Richard A. Carrigan, Jr., and James A. Ellison. NATO ASI Series, Series B: Physics, Vol. 165. Plenum Press, 1987.

CHANNELING AND OTHER PENETRATION EFFECTS. Edited by E. Uggerhøj. Special volume of Nuclear Science Applications, in press.

Caprice Classic Brougham. Now you can comfortably afford to ride in the lap of luxury.

- Standard V8 with Electronic Fuel Injection.
- 4-speed automatic overdrive transmission.
- Standard AM/FM stereo sound system.
- Available genuine leather seating areas for six.
- New 3-year/50,000-mile Bumper to Bumper Plus Warranty.*

THE *Heartbeat* OF AMERICA TODAY'S CHEVROLET™

You see here two different ways to travel in the lap of luxury. The one in front is ours. Caprice Classic Brougham is so wondrously smooth, you sometimes forget you're moving. So comfortable, you'll never want it to stop. And so affordable, you'll probably be able to take a ride in the other luxury vehicle shown here, as well.

*See your Chevrolet dealer for terms of this new limited warranty.

Authenticating Ancient Marble Sculpture

Stone sculpture is notoriously difficult to prove genuine. As prices for such works skyrocket, geochemists are being called in to help settle questions that art history and a connoisseur's eye cannot resolve

by Stanley V. Margolis

In 1984 the J. Paul Getty Museum in Malibu, Calif., was offered a chance to buy an archaic Greek kouros: a larger-than-life marble figure of a youth. The two-meter-tall statue, said to be more than 2,500 years old, was superbly preserved but had been unknown to art historians; it was said to have resided in a private Swiss collection for many decades. Based on the putative importance of the piece, the owner was asking a higher price than had ever been paid for an ancient statue; several major newspapers reported that the sum was somewhere between eight and 12 million dollars.

According to Marion True, the museum's curator of antiquities, most of the art historians who had examined the kouros judged it to be authentic. Others, however, considered it "questionable." They were particularly troubled by its remarkable preservation and by certain stylistic anomalies. In addition, the surface had suspicious characteristics when examined under ultraviolet light. The conventional wisdom—not proved, however, by rigorous experiment—is that ancient marble surfaces fluoresce an amber color mottled with purple. The kouros fluoresced an uneven light purple, a color associated with modern surfaces. Given the statue's potential importance as well as its price and the doubts that had been raised, museum officials wanted conclusive scientific evidence for its authenticity before they would consider buying the piece.

Stone objects are much more difficult to authenticate than paintings, ceramics and other human artifacts. The stone's absolute age reveals nothing about when it came under the sculptor's chisel. Methods do exist, however, for determining when a stone surface was first exposed to the environment. Geochemists and geologists routinely date the age of exposed rock outcrops on the basis of the weathering that occurs when the rock surface interacts with its chemical environment. Nevertheless, curators and collectors as a rule had not consulted geologists about the authenticity of marble sculpture; they no doubt had fearful visions of scientists attacking their priceless masterpieces with hammers and drills.

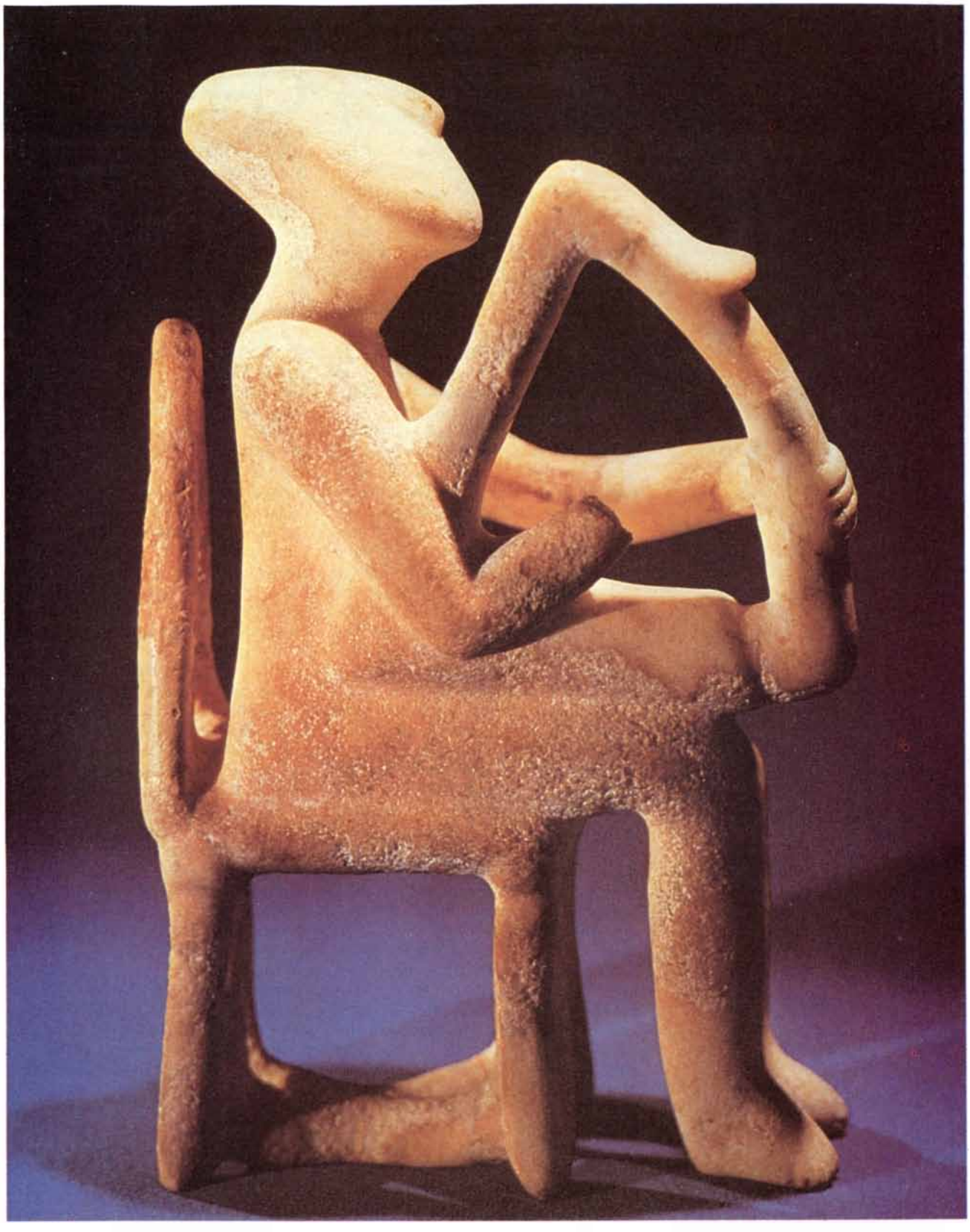
It was therefore a new experience for everyone when the Getty Museum invited me to apply my skills as a marine geochemist in determining the age of the kouros. I was even given permission to remove and analyze small samples of the stone. By the time I had finished the analysis, more than a year later, the kouros had probably undergone the most complete scientific examination of any ancient marble sculpture to date; it was certainly the first time a museum had followed such a procedure to make a procurement decision. In part because of this work, scientific methods for authenticating marble sculpture are now widely accepted by curators at major museums around the world.

In theory many kinds of stone artifacts can be dated by geochemical methods, but marble has a special importance. For more than 7,000 years, marble sculpture has mirrored the artistic and spiritual progress of Western civilization. Sculptors have prized the translucent surface of white marble and its ability to accept a fine polish. Early Bronze Age inhabitants of the Greek Cyclades islands carved exquisitely simple figurines from marble cobbles found on the ground. By the first millennium B.C. the Greeks were quarrying large marble blocks on the island of Naxos, probably using techniques learned from the Egyptians. During Roman times Greek sculpture was highly coveted and widely copied. Collectors from the Renaissance to the present have lavished princely sums on marble statuary from the ancient world.

For about as long as there has been a market for marble sculpture, there has also been a thriving industry in copies and forgeries. To tell authentic pieces from fakes, experts have relied on each object's style, iconography and art-historical context. They have also relied on the appearance of the surface layer, or patina, to guess the age of the carving. Marble is quite resistant to weathering under most natural conditions, however, and the effects of aging and exposure are often invisible to the unaided eye. What is more, forgers often try to fake a patina by a variety of methods, from burying in cow pastures to applying acid or pastes that mimic an ancient crust. Many controversies over important marble pieces have remained unresolved for lack of objective evidence.

In the course of their own investiga-

STANLEY V. MARGOLIS is professor of geology at the University of California, Davis, and adjunct professor of marine science at the University of California, Santa Barbara. He received his B.A. in marine science and geology from the University of Miami and his Ph.D. in geology and oceanography in 1971 from the University of California, Riverside and Santa Barbara. In 1981 he served as head of ocean sciences for the Naval Ocean Research and Development Activity. By analyzing marine sediments, Margolis has investigated the history of polar glaciations and global climatic changes and the environmental effects of ocean mineral exploitation. He is currently helping to form a research consortium in California for art-conservation science and geoarchaeology.



HARPIST of calcitic marble in the Virginia Museum of Fine Arts is thought to have been carved 4,600 years ago in the Cyclades islands of Greece. Its authenticity is debated because of stylistic anomalies in the chair and the placement of the harp. The brown patina is flecked with unusual white deposits of gypsum. Isotopic analysis showed that the marble probably came from outcrops on the island of Naxos, a source consistent with the piece's putative origin. The gypsum could have been deposited by the evaporation of saline groundwater, which may explain why the crust is enriched in carbon 13 and oxygen 18. An ongoing survey of Cycladic sculpture should help determine whether these geochemical features are typical.

tions, meanwhile, geochemists have learned a great deal about marble and its precursors: carbonate sediments and rocks such as limestone, which undergo metamorphism into marble when subjected to high temperature and pressure. Such rocks are studied to reconstruct the history of the earth's changing climate and environment. Analyses of carbonates drilled from the seafloor, for example, have revealed the timing and origin of the ice ages and polar glaciations. Studies of carbonate rocks from the boundary between the Cretaceous and Tertiary periods—the time about 65 million years ago when the dinosaurs and many other plant and animal species became extinct—can help to reconstruct the environmental conditions associated with the extinction.

Carbonate minerals can be studied for clues not only to the history of the earth but also to the origin of the rock sample itself. The proportions of various stable carbon and oxygen isotopes in limestone and marble vary depending on the specimen's geologic and geographic provenance. Isotopic analysis also provides clues to the structural and compositional changes caused by weathering and burial. The method is augmented by various other tools. Such tools include polarized-light microscopy, which shows alterations in the mineral structure, and electron microprobes, which stimulate rock samples with high-energy electron beams; the frequency of the X rays the samples emit in response is analyzed to determine trace-element concentrations.

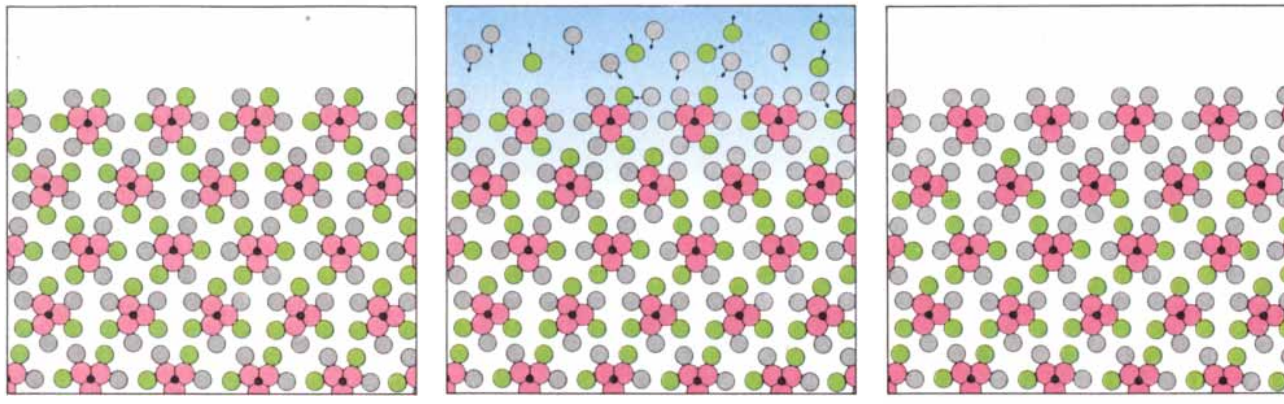
The same instruments and scientific principles can be applied to marble sculpture. Although geologic weathering takes place over millions of years, whereas the weathering of ancient marble sculpture occurs over a few thousand years, the basic processes are the same. What is more, such changes progress faster on land than in the ocean, which is an extraordinarily stable environment. Terrestrial environments are more variable, and rain and groundwater are highly corrosive to carbonate rocks. Indeed, a detailed understanding of weathering in marble sculpture could contribute to the study of changes in carbonate rocks that may accompany such processes as climatic change, which occur over short time scales.

Our foray into the world of art began in the summer of 1984. The Getty Museum had begun a project to safeguard its large collection of 17th- and 18th-century French and Italian marbles against the earthquakes that frequently shake Malibu. Small holes were drilled in the marble objects so that anchor cables could be attached. The museum decided to have the drilled cores analyzed in order to learn more about the objects. I and my wife Karen, a scanning electron microscopist, were asked to identify the marble types as well as their origin and history. This was the first time modern petrographic techniques were employed to study decorative marbles. We must have presented an unusual sight with our high-tech equipment amid busts of Louis XIV and Marie Antoinette's commode.

One afternoon Jerry Podany, the antiquities conservator at the museum, joined us for lunch, and a stimulating conversation ensued about scientific investigations of marble authenticity. Afterward we were taken down to the museum basement and shown an ancient kouros, lying covered with a cloth on a cushioned cart in a small locked room. It was in seven pieces, the largest consisting of the head and torso. The statue appeared to be complete and in excellent condition; only the tips of the nose and penis were missing. Five of the fragments seemed to have broken in ancient times; the broken surfaces were covered with a tannish-red patina that resembled the patina on the sculpted surface. Two of the fractures—the ones detaching the right arm and hand—appeared to be recent, since the broken surfaces were snow-white.

Podany told us the museum was considering buying the kouros, and he wanted to know if there were any chemical or mineralogical tests that could clarify the age and history of the statue. We replied that the statue would have to be carefully examined with a microscope and samples of the stone analyzed before we could say anything definitive. For the next two days we examined the entire surface under a high-resolution stereomicro-

ANCIENT GREEK KOUROS, thought to have been carved between 540 and 520 B.C., is shown here reassembled from seven fragments. When the statue was offered for sale to the J. Paul Getty Museum, doubts were raised because of its superb condition and eclectic style, which seems to combine features from different periods and regions. The statue is carved from dolomite, a highly stable type of marble; this may explain why it has only a very thin patina. The kouros stands 206 centimeters high.



DEDOLOMITIZATION is depicted schematically. Dolomitic marble is composed of crystals of calcium magnesium carbonate (*left*). When the surface is exposed to groundwater containing large amounts of dissolved calcium ions, the calcium gradually replaces the magnesium in the dolomite (*middle*). After many centuries the exposed surface of the dolomite has changed into calcite, or calcium carbonate (*right*).

scope, much as we would when conducting a geologic "autopsy" of natural rocks. After the initial examination, we told Podany and John Walsh, Jr., director of the museum, that it would be necessary to drill a small core. A few days later the owner of the kouros gave permission for the sample to be taken. After dry runs on some marble scraps, Podany and his staff, not without some anxiety, carefully drilled a core measuring one centimeter in diameter and two centimeters in length from the surface of an ancient fracture below the right knee.

In our study of the core only one question concerned us: Could the surface have been sculpted between 540 and 520 B.C., as suggested by art historians, or had it been carved in more recent times and then treated artificially to appear ancient? To begin we split the core lengthwise and then sliced one part into thin, longitudinal sections, which yielded profiles of the outer weathered surface, the patina crust and the transition to the fresh marble in the interior. The sections were examined by electron microscope and electron microprobe. Other samples were subjected to mass spectrometry, X-ray diffraction and X-ray fluorescence to determine the abundance of various isotopes, minerals and trace elements.

The initial examinations revealed that the marble was composed of relatively pure dolomite (calcium magnesium carbonate), a variety of marble that is less common than those composed of calcite (calcium carbonate). Dolomitic marble is more durable and generally more resistant to weathering than calcitic marble, and that could explain the excellent preservation of the kouros.

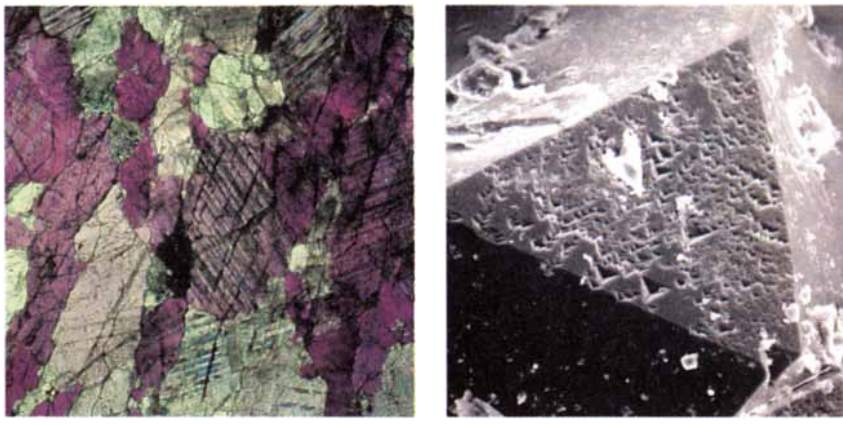
Usually the first step in authenticating a work of sculpture is to identify the quarry from which the stone came. This can often be done with great precision, because the isotopic makeup of marble varies significantly from one site to another. When isotope values overlap from one quarry to another, further chemical tests can tell the quarries apart. The kouros marble was analyzed by our research group, which included William S. Showers of North Carolina State University at Raleigh and Norman Herz of the University of Georgia, the latter an expert in the provenance of ancient marble. According to Herz, the kouros marble probably came from the ancient Cape Vathy quarries on the island of Thasos. These quarries were the oldest source of dolomitic marble in the area, and large kouroi were produced in Thasos in the seventh and sixth centuries B.C.

Although the analysis indicated a reasonable provenance and historical context for the kouros, it did not prove its age or authenticity, since Thasian marble is still available. Even if the marble happened no longer to be available, the statue could still have been reworked from blocks taken from an archaeological site. The question of authenticity, then, had to be answered by determining the age of the sculpted surface.

Under an optical microscope the kouros surface revealed a thin, tan patina of iron oxides and soil-clay minerals, along with manganese-oxide encrustations similar to those found on ancient marble outcrops. In nature these minerals originate from soil and from groundwater. When the core section was examined more closely with the electron microprobe, the weathered surface was found to be covered with a continuous layer of calcite, about 10 to 50 microns thick. The identification of calcite was confirmed by both X-ray diffraction and electron-microprobe analysis, carried out by me and Peter Schiffman of the University of California at Davis and independently by Donald Dietrich and Frank Preusser at the Getty Conservation Institute in Marina del Rey.

The finding of a calcite layer on the kouros was significant because dolomite is well known to change into calcite through weathering. In this process, known as dedolomitization, magnesium atoms in dolomite crystals are slowly replaced by calcium atoms. Calcite crusts are often found on portions of ancient dolomitic outcrops that have been exposed to low-salinity groundwater or rainwater that is enriched in calcium and poor in magnesium. On the kouros, calcite appeared on fractured surfaces that looked weathered but not on any of the fresh-looking ones. Presumably the fractured surfaces that have a calcite layer had been exposed to weathering in ancient times.

We next applied softened acetate tape to several sculpted surfaces and obtained detailed "fingerprints" of surface features. The tape also picked up mineral grains from the surface crusts without visibly damaging the statue. The tape was examined by scanning electron microscope and electron microprobe. All of the samples taken from the sculpted surface had the same calcite crust, clay minerals, trace elements and iron oxides as



KOUROS DOLOMITE'S distinctive crystal structure, shown in a polarized-light micrograph (*left*), is consistent with dolomite from Thasos. A scanning electron micrograph of a core sample (*right*) shows triangular pits etched by corrosive groundwater over a long time on the face of a dolomite crystal (enlarged 300 diameters).

the cores from the weathered fractured surfaces did. Evidently the calcite layer extended over the entire weathered surface of the kouros.

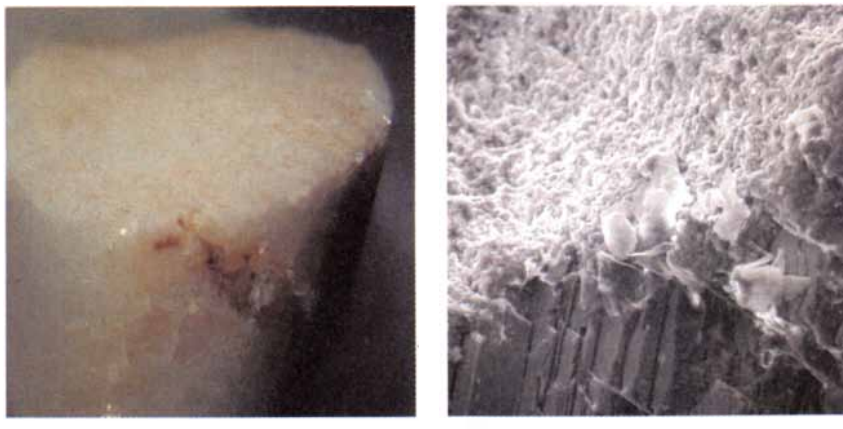
Evidence gathered in the field and from laboratory experiments indicates that under normal conditions at the earth's surface dedolomitization requires long time spans, from hundreds to thousands of years. Attempts to accelerate the process, for example, by boiling dolomite grains at 350 degrees Celsius under 250 atmospheres of pressure for almost one year, succeeded only in changing a tiny fraction of the dolomite into calcite and brucite (magnesium hydroxide), a result that does not resemble the effect of dedolomitization.

Given that specks of dolomite can be converted to calcite only with great difficulty in a modern laboratory, it seemed inconceivable that the calcite on the kouros was produced artificially in a uniform layer over the two-meter length of the statue. It also seemed unlikely that an artificial calcite layer was applied to the stone. The strontium, manganese and other trace elements found in the fresh dolomite and in the calcite crust were similar, which would indicate that the calcite developed in situ.

In the most crucial test we compared the ratio of oxygen 16 to oxygen 18 and that of carbon 12 to carbon 13 found in the calcite crust with the ratios found in the fresh dolomite from the interior. The ratios were similar and also resembled the ratios measured in dedolomitized dolomite found in nature. This finding ruled out the possibility that the calcite crust of the kouros had been artificially replicated or had precipitated out of groundwater. In either case the oxygen and carbon isotope values of the crust would have been very different from those of fresh dolomite. The most reasonable explanation for the calcite on the kouros is that it developed through centuries of weathering.

The thickness of the calcite crust by itself does not indicate an exact age for the sculpted surface, because the rate of dedolomitization depends on many unknown variables: the length of burial, the composition of the burial soil and the chemistry of groundwater and rainfall at the burial site. From what is known about dedolomitization, the surface features of the Getty kouros are consistent with an age of no less than several centuries and no more than several millennia.

Our geochemical findings underwent international scientific review and verification in blind studies by independent laboratories. In addition, the curatorial staff of the Getty Museum carried out a detailed stylistic analysis and compared the statue with 200 other complete or fragmentary kouroi. After 14 months of intensive study the museum decided to purchase the kouros. In the fall of 1986 the statue went on display for the first time. Held together internally by a novel, earthquake-proof system of stainless-steel cables and high-tension springs, it now stands complete and upright, appearing much as it must have in ancient times.



CORE SAMPLE the size of a pencil eraser (*left*) was drilled from a weathered fracture at the right knee of the kouros. Reddish iron oxide stains the calcite crust on the exposed surface and penetrates deeply into a fissure; the fresh dolomite is snow-white. A scanning electron micrograph (*right*) shows the tight bonding of the calcite crust to the underlying columns of dolomite crystals, here enlarged 1,500 diameters.

The methods employed on the Getty kouros cannot be applied universally to authenticate all marble sculpture. They are relevant only to dolomitic marble. Calcitic marbles used in ancient sculpture, such as Parian, Doliana and Naxian marbles, are far more common than dolomitic marbles. They also develop a weathering layer, but in a different manner. As a result they require somewhat different analytical techniques and criteria for authentication.

When a sculptor works a piece of marble, the hammerblows "stun," or fracture, a thin surface layer of crystals. These fractured crystals allow water to penetrate. Calcite is more soluble than dolomite, so that in sculpted calcite surfaces the percolation of water causes large calcite crystals to recrystallize into a much finer structure called micrite (microcrystalline calcite). This micrite layer is opaque, and it thickens as it is exposed to the elements and to percolating groundwater.

In weathered calcite the crystal structure changes gradually from micritic calcite at the surface to fresh

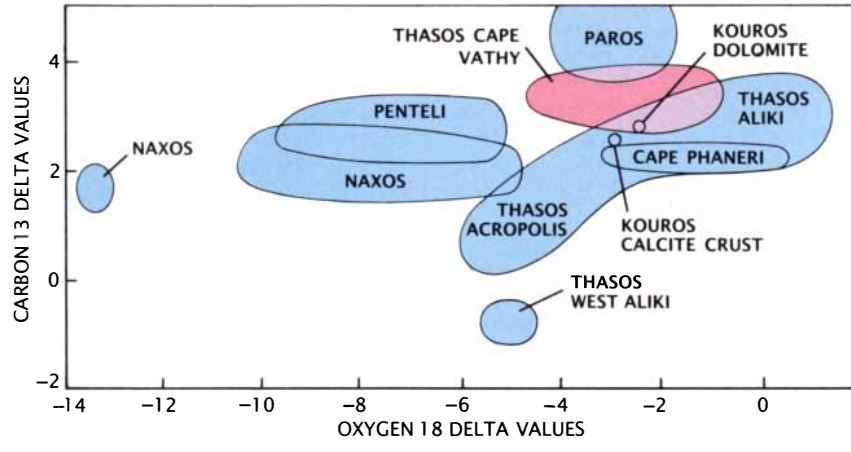
marble in the interior. The crust can vary in thickness from 10 microns to more than 10 millimeters, depending on the length of exposure and the composition of the groundwater. The exact rate of weathering also depends on such characteristics of the marble as the crystal size and intergranular porosity. The biological activity of lichens and algae on marble surfaces may also play an important role in crust formation.

Isotopic analysis shows that there is a progressive increase in the abundance of the lighter carbon and oxygen isotopes carbon 12 and oxygen 16 from the fresh interior marble to the weathered crust. These isotopes are more abundant in the carbon dioxide dissolved in groundwater than they are in calcite, so that as the calcite undergoes micritization the micrite becomes enriched in them. The rate of enrichment varies depending on the thickness of the existing crust and the amount of various isotopes and minerals present in the groundwater. Authentic ancient marble sculpture consistently shows a difference in the abundance of carbon and oxygen isotopes between the fresh and the weathered marble, whereas forgeries and modern works do not.

Clay minerals and trace elements in the soil such as iron and manganese are also incorporated in the crust. These substances permeate ancient marble well below the surface and are not just loose surface impurities, as would be the case in a forgery. Another sign of antiquity is the presence of calcified plant rootlets and lichens. In addition, over a long period naturally acidic water etches delicate patterns in the surface; attempts to mimic the patterns by pouring acid on marble will appear obviously different.

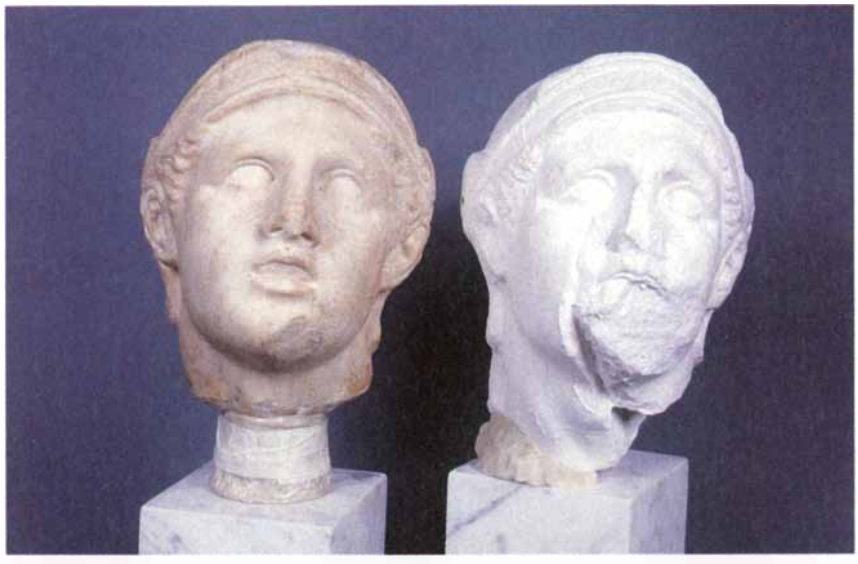
Until recently restorers cleaned marble sculpture with acid, an unfortunate practice that removed the weathering crust and made it difficult to apply these methods of authentication. Indeed, the authenticity of many works will remain in question forever because they were acid-cleaned. Modern conservators generally condemn such practices.

"MARBLE ISLANDS" of Naxos, Thasos and Paros are shown on this map of the Aegean region. Archaic and classical Greece and ancient Rome quarried marble from these islands, as well as from the Greek mainland, Anatolia and Italy. During the Mesozoic era (between 225 and 65 million years ago), the area was submerged under tropical seas. Skeletons of sea fauna formed thick carbonate deposits on the seafloor. Over time the deposits were metamorphosed and uplifted to form beds of white marble.



ISOTOPIC ANALYSIS provided key evidence for the authenticity of the Getty kouros. The graph indicates the amount of oxygen 18 and carbon 13 in various types of calcite (*blue*) and dolomite (*red*) as compared with a standardized sample; the units are delta values, or divergences from the standard in increments of .1 percent. Lower delta values indicate correspondingly higher concentrations of the lighter isotopes oxygen 16 and carbon 12. The analysis shows that the kouros dolomite probably came from the Cape Vathy quarries on Thasos. The calcite crust is slightly enriched in oxygen 16 and carbon 12; because these isotopes are more abundant in groundwater than in dolomite and would gradually replace heavier isotopes in the stone, it is likely that the crust developed through natural weathering. The data on isotope values for the quarries were supplied by Norman Herz of the University of Georgia.

These techniques for authenticating calcitic marble sculpture were employed recently on several questionable pieces. A head of Achilles that had once been displayed at the Getty Museum came under suspicion. The head was ascribed to the fourth-century B.C. Greek sculptor Skopas, and it closely resembled a head of known authenticity in the Na-



HEAD OF ACHILLES owned by the Getty Museum (*left*) was compared with a head in the National Archaeological Museum in Athens, shown here in a plaster cast (*right*). Both supposedly came from a temple at Tegea, but isotopic analysis revealed that the Getty head was carved from Parian marble, whereas the one in Athens was probably of Doliana marble. Also, the Getty head lacked a weathering layer and other signs of antiquity. The evidence suggests that it is a 20th-century copy.

tional Archaeological Museum in Athens that was thought to have come from the same Greek temple. The Getty Museum found stylistic inconsistencies in its own sculpture, however, and measurements showed that the two heads were virtually identical in all dimensions. The evidence suggested that the Getty head had been copied from the Athenian one.

Marble samples were taken from the two heads and analyzed for oxygen and carbon isotopes at two independent laboratories. The tests revealed that the two marbles were quite different in type, which made it even less likely that the Getty head was genuine. In addition, the samples taken from the Getty head showed that the surface lacked a weathering layer as well as all the other geochemical and petrographic features one would expect to see in ancient marble sculpture of that period. Together these findings indicated that the Getty head is probably a 20th-century copy.

Another case involved a Greek Cycladic figure of a seated harpist in the Virginia Museum of Fine Arts in Richmond. The figure had been ascribed to the third millennium B.C. but came to be considered questionable because of stylistic quirks and also because it bears an unusual weathering crust containing traces of the mineral gypsum (calcium sulfate). Margaret Mayo, the museum's curator of ancient art, asked our laboratory to examine and analyze the piece.

We found that the harpist was carved from calcitic marble that probably came from the island of Naxos, which was the most likely source of the marble in the third millennium B.C. It had a thick weathering crust infiltrated by trace elements in a way that was consistent with an age of 5,000 years. Carbon 13 and oxygen 18, on the other hand, were more abundant near the surface than in the interior. This is the reverse of what we had observed previously in ancient calcitic marbles, in which the lighter isotopes carbon 12 and oxygen 16 are more abundant near the surface. We know, however, that in arid climates gypsum and calcite deposits can form on surfaces through evaporation of saline groundwater and other moisture. Such evaporites could have incorporated the heavier isotopes into the crust.

Other Cycladic sculptures are now being examined and compared in detail by a group of scientists from several U.S. museums, including the Boston Museum of Fine Arts, the Metropolitan Museum of Art in New York, the Los Angeles County Museum of Art and the Smithsonian Institution. A verdict on the authenticity of the Virginia Museum's seated harpist must await the results of this survey.

In the past four years we have made much progress toward comprehending the natural weathering of ancient marble sculpture. Our knowledge base expands with each piece that comes under scientific scrutiny. Techniques and instruments have also improved, so that analysis can now be based on samples smaller than the head of a pin, taken from areas hidden from the view of even the most hawk-eyed connoisseurs.

Curators now see the value of scientific techniques not only for authenticating marble sculptures but also for dating and conserving them. Several large museums in Europe and the U.S. are now carrying out major geochemical surveys of marble sculpture. In the course of these surveys, investigators have found other archaic Greek statues of Thasian dolomite that have calcite patinas similar to the one seen on the Getty kouros. They have uncovered some forgeries and, in other cases, have proved the authenticity of objects that had been in doubt. Even the most skeptical critics now agree that scientific methods should be added to art-historical and archaeological analyses. Indeed, many curators who once doubted the methods applied to the Getty kouros can now expound on dedolomitization.

The science of authenticating ancient marble sculpture is still in its infancy. It will be refined as more examples of both questionable sculpture and sculpture of proved authenticity are analyzed, together with samples of raw marbles of various ages and types from both ancient and modern quarries. Sophisticated new techniques based on stable and radioactive isotopes may make it possible to determine the absolute age of weathering crusts. These studies are needed not only to advance this new field of science but also to ensure that science remains one step ahead of those who, for private gain, would falsify the artistic legacy of human history.

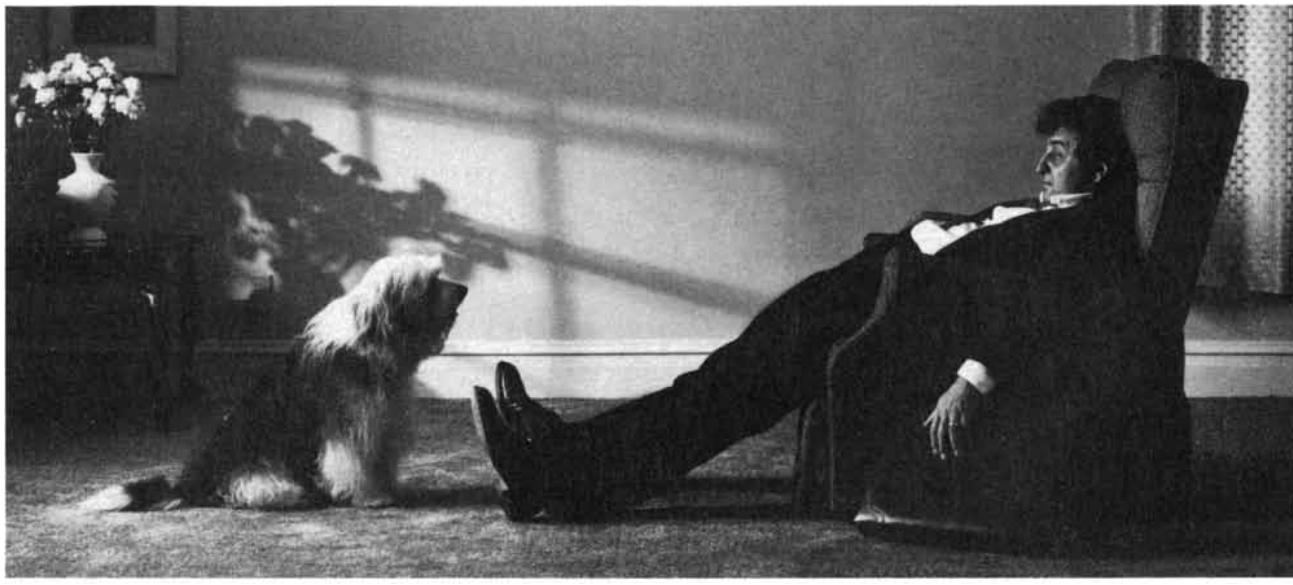
FURTHER READING

ISOTOPIC ANALYSIS OF MARBLE. Norman Herz in *Archaeological Geology,* edited by George Rapp, Jr., and John A. Gifford. Yale University Press, 1985.

A KOUROS AT THE GETTY MUSEUM. Marion True in *The Burlington Magazine,* Vol. 129, No. 1006, pages 3–11; January, 1987.

WEATHERING CHARACTERISTICS, AGE, AND PROVENANCE DETERMINATIONS ON ANCIENT GREEK AND ROMAN MARBLE ARTIFACTS. Stanley V. Margolis and William Showers in *Classical Marble,* edited by N. Herz and M. Waelkens. Kluwer Academic Publishers, 1988.

THE GETTY MUSEUM KOUROS: ANALYSIS, AUTHENTICATION AND CONSERVATION. Jerry Podany et al. J. Paul Getty Museum Press, in press.



How to solve the energy crisis.

If you're like a lot of people the energy crisis has nothing to do with oil embargos and solar power.

You're suffering from a personal energy shortage. You know, no pep, no get-up-and-go, no spunk.

The reason for this phenomenon? Lack of exercise.

Exercise equals energy.

Research has proven that people who exercise on a regular basis have more stamina and feel better about themselves.

And health and fitness experts agree that no form of exercise is more efficient and effective than cross-country skiing. Not walking, running, rowing, or biking.

NordicTrack® gets you back on the right track.

NordicTrack simply duplicates the cross-country skiing motion, but you don't have to know how to cross-country ski to use it.

This no-impact, total body workout provides a better cardiovascular workout than exercise bikes, rowers, and treadmills.

You burn more calories. You have more energy.

With NordicTrack, you can burn more calories in a 20-minute workout than with any other type of exercise machine.

Plus, you'll feel stronger, more alert, and less stressed-out.

Why NordicTracks never show up at garage sales.

People love their NordicTracks. In fact, 7 out of 10 owners are still using their machines more than 3 times a week, 5 years after purchasing one.

So call NordicTrack.

But do it soon. Before you run out of gas for good.

Free Brochure & Video.
Call Toll Free 1-800-328-5888.
In Canada 1-800-433-9582.

☐ Please send me a free brochure.
☐ Also a free video tape ☐ VHS ☐ BETA

Name_____
Street_____
City_____ State_____ Zip_____
Phone ( )_____

NordicTrack 141 Jonathan Blvd. N. Chaska, MN 55318
320F9

NordicTrack
A CML COMPANY

Absinthe

Evidence of the pale-green liqueur's toxicity eventually extinguished the fin-de-siècle infatuation with absinthe. The drink's history began, however, long before the 19th century

by Wilfred Niels Arnold

Vincent van Gogh shot himself on the afternoon of July 27, 1890, in Auvers-sur-Oise, France; he died in the early morning two days later. Paul F. Gachet, the doctor who attended van Gogh during the last two months of his life, planted a thuja tree on the artist's grave. The gesture was probably inspired by van Gogh's admiration of thuja trees and his inclusion of their flamelike images in some of his Auvers paintings.

Gachet's choice of a grave ornament was unwittingly pathetic. The thuja tree is a classical source of the chemical thujone, a constituent and indeed the toxic principle of the alcoholic drink known as absinthe. There is good evidence to indicate that van Gogh was addicted to absinthe, that his psychosis was exacerbated by thujone and that his fits with hallucinations contributed to his suicide.

In his fondness for absinthe van Gogh was by no means alone. The drink was enormously popular in the late 19th century, particularly in France. French soldiers fighting in the Algerian conflicts of the 1840's had spiked their wine with wormwood extract (ostensibly to ward off fevers), and on their return to France their acquired taste was satisfied by absinthe, which contained a variety of essential oils including that of wormwood. Absinthe's popularity with the soldiers spread among their compatriots from all walks of life; some of the most creative people of the time were its devotees. Absinthe was said to evoke new views, different experiences and unique feelings.

It could also wrack the drinkers' brains. The disease known as absinthism was recognized in the 1850's; its victims evinced a dazed condition and intellectual enfeeblement and experienced terrifying hallucinations. The symptoms and extent of the damage from excessive consumption of absinthe could not be attributed to alcohol alone. Other culpable chemicals came from the leaves and flowers used in the drink's preparation. But manufacturers, governments and the public, enamored of profits, tax revenues and titillation, respectively, were slow to heed the warning signs. Absinthe was not banned until the 20th century.

There may have been a subtler reason for the reluctance to abandon this favorite spirit. Some of the plants that gave absinthe its distinctive taste were stock remedies from herbal lore; they had been exploited for thousands of years with results that were often meritorious, sometimes innocuous and rarely sinister. Even after the liqueur's fall from grace, investigations of the chemistry and the physiological effects of its constituents, as well as those of related chemicals, contributed to medical practices and to the development of effective drugs.

WILFRED NIELS ARNOLD is professor of biochemistry at the University of Kansas Medical Center. Born in Brisbane, Australia, he has a B.Sc. from the University of Queensland, an M.A. from the University of California, Los Angeles, and a Ph.D. from Cornell University, awarded in 1962. His biochemical research centers on yeasts, but his avocational interests, which led him to write this article, include the history of 19th-century medicine and art history.

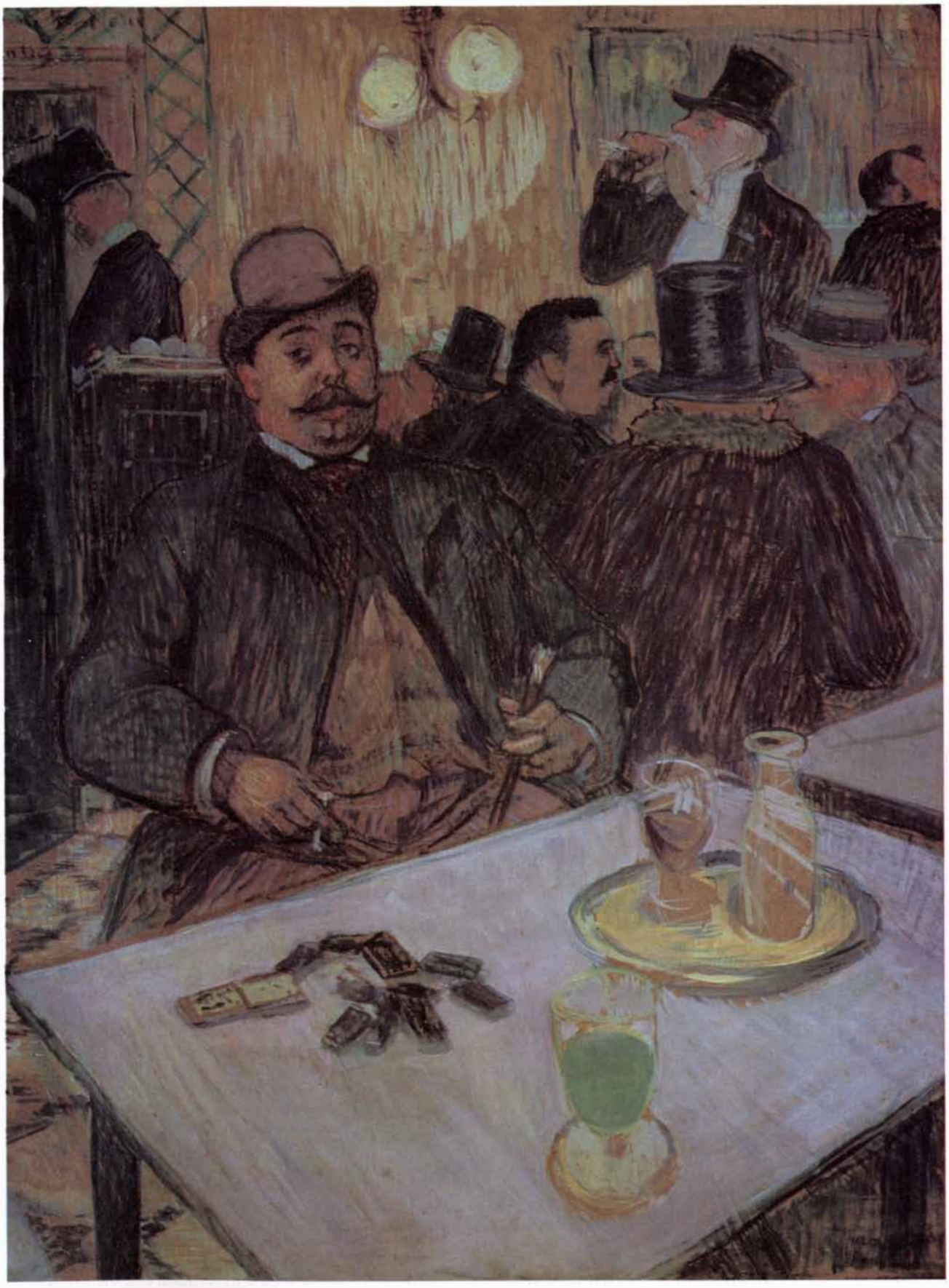
Thujone occurs in a variety of plants, including tansy (*Tanacetum vulgare*) and sage (*Salvia officinalis*), as well as in all trees of the arborvitae group, of which the thuja (*Thuja occidentalis*), or white cedar, is one. It is also characteristic of most species of *Artemisia,* a genus within the Compositae, or daisy, family. Wormwood (*Artemisia absinthium*) and Roman wormwood (*Artemisia pontica*) were the main sources of the thujone in absinthe.

Wormwood (in French, *absinthe;* in German, *Wermut*) is an herb with a perennial root system from which arise branched, firm, leafy stems that are almost woody at the base and reach a height of two to three feet. Its flowers are tiny, greenish-yellow and globular, and its indented leaves have a silvery-gray sheen [*see illustration on page 114*]. The species was cultivated from the Middle Ages to the early part of the 20th century.

The earliest recorded use of wormwood comes from the Ebers Papyrus, copies of which date from 1550 B.C. but which include writings from 3550 B.C. To the Egyptians, wormwood, or a closely related species, had religious as well as medicinal significance. The "wormwood" that is mentioned seven times in the King James' version of the Bible was probably not *Artemisia absinthium,* but *Artemisia judaica.* Pliny's *Historia Naturalis,* written in the first century A.D., describes extracts of wormwood as being of great antiquity (even then!) and having long-standing utility against gastrointestinal worms (hence the name). Thujone does indeed stun roundworms, which are then expelled by normal peristaltic action of the intestine.

Wormwood was fully described in Dioscorides' *De Materia Medica,* an influential book that was completed in about A.D. 65 and was considered the final authority in pharmacy for the next 1,500 years. Both Pliny and Dioscorides included several applications for wormwood in addition to its anthelmintic properties. Anointing arms and legs with the plant's juice helped to repel gnats and fleas, and attaching leaves to stored garments protected them from moths. These "virtues" have been substantiated, but the authors go on to list others that have not.

Pliny also mentioned a wine known as absinthites that was fortified with extract of wormwood. From the first to the 15th century, however, the selection of wormwood, tansy and other plants as additives to foods and beverages was supposedly based on



PAINTINGS OF CAFÉ LIFE in turn-of-the-century France often represent absinthe as part of the social scene. A glass of absinthe is conspicuous in this detail of the 1893 painting by Henri de Toulouse-Lautrec, *Monsieur Boileau at the Café.*

their characteristic taste rather than on their ability to intoxicate. With the development of steam distillation in the 16th century (described in the books Hieronymus Brunschwig published in 1500 and 1512), relatively innocuous decoctions were replaced by powerful essences derived from the same plants. In the 17th century, tansy became popular in a baked dish of the same name that was made with eggs and cream. *Artemisia maritima* was used to make purl, a fortified ale that was popular in 17th- and 18th-century Ireland and England. Purl is mentioned in *The Merry Wives of Windsor,* and wormwood is mentioned in three other Shakespearean plays.

WORMWOOD, *Artemisia absinthium,* is shown in James Sowerby's 1803 hand-colored engraving. Oil of wormwood was extracted from the herb's leaves, flowers and stem; it gave absinthe a distinctively bitter taste. Wormwood oil includes thujone, which can cause hallucinations, convulsions and permanent damage to the nervous system.

The production of grain alcohol by distillation of fermented cereals heralded the invention of liqueurs, and toward the end of the 18th century the formulation of absinthe evolved in Switzerland. The recipe found its way to Henri-Louis Pernod, who in the early 1800's opened a factory in Pontarlier, France. Several competitive companies were subsequently founded in France and Switzerland, and for the next 100 years absinthe production was a significant industry.

Rue is said to be the most bitter plant known, but wormwood is a close second. The bitterness is due to a compound called absinthin ($C_{30}H_{40}O_6$), the complex structure of which was not solved until the 1950's. The bitterness threshold for pure absinthin is one part in 70,000: one ounce can be detected in 524 gallons of water.

To overcome the bitter taste of absinthe, it was customary to add a sweetener. The most genteel manner involved mounting a cube of sugar in a silver sieve (an absinthe spoon) that was placed across the top of a glass containing a small amount of absinthe. Cold water was then poured over the sugar cube into the glass. Dilution turned the clear green of the liqueur to a yellow opalescence. Men and women became enthralled with this ritual of presentation as well as with the appearance, taste and excitement of the liqueur.

The aesthetics of absinthe drinking may account in part for the aura that soon enveloped it. In the cheerful atmosphere of recovery that followed the Franco-Prussian War (1870–1871), *l'heure verte* ("the green hour") became an established daily event; some Parisian clubs and cafés were dedicated to the liqueur. Images of absinthe are immortalized in paintings such as Édouard Manet's *The Absinthe Drinker* (1859), Edgar Degas' *L'Absinthe* (1876) and Henri de Toulouse-Lautrec's pastel of van Gogh with a glass of absinthe, which was completed in 1887. The same year van Gogh himself did a still life with a glass of absinthe and a carafe.

The incisive graphic work of Honoré Daumier addressed the subject with social commentaries such as the lithographs entitled "Beer—never . . . it takes absinthe to revive a man" and "Absinthe . . . the first glass . . . the sixth glass," which were published in *Le Charivari* in 1863. *Absinthe Drinkers,* an 1881 canvas by Jean-François Raffaelli, has subjects that are more mellow than depressed, and the two glasses are truly opalescent. Pablo Picasso created *Absinthe Drinker* in 1901 and *The Poet Cornuty—Absinthe* in 1903. Eleven years later he constructed six abstract glasses of metal and ceramic topped with absinthe spoons—one artist's response to increasing legislative attempts to ban absinthe in France.

Charles Baudelaire, poet and a close friend of Manet, included absinthe in his list of vices; he advised, "Be drunk, always," but went on to say, some lines later, "With wine, with poetry, or with virtue, as you please," which gives a more wholesome choice than is typically attributed to the author. Paul Verlaine often awaited inspiration over a glass and then wrote in tones rampant, coarse and sensual within the same verse. Arthur Rimbaud, whose brilliant poetic career was finished by the time he was 20, was in his cups for most of that brief time.

The English poet Ernest Dowson punned on an aphrodisiacal rumor: "I understand that absinthe makes the tart grow fonder." The enigmatic surrealist playwright Alfred Jarry insisted that rational intelligence was inferior to hallucinations and relied on absinthe to ensure a steady supply of the latter. In the bistros of the rue de Seine, Guillaume Apollinaire, poet and friend of Picasso and Gertrude Stein, came under the spell of both Jarry and the ever-present absinthe.

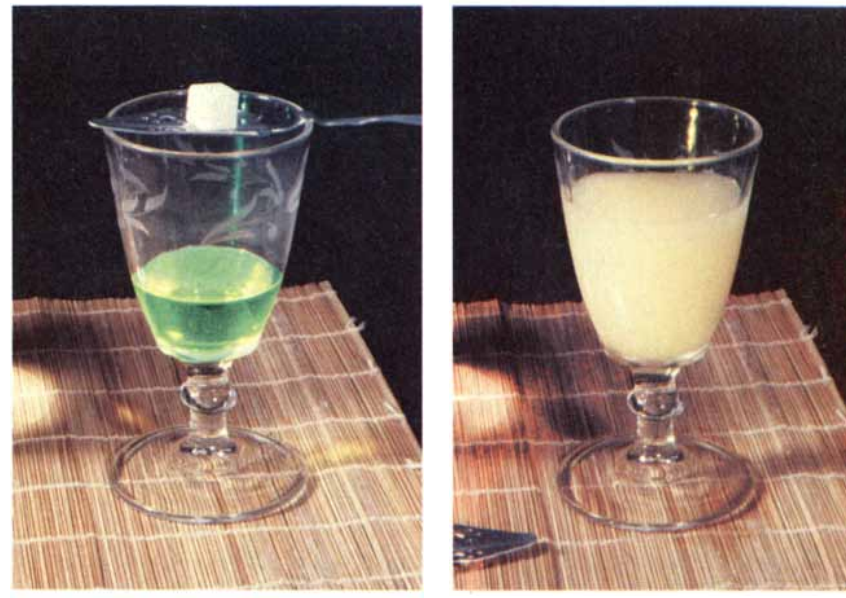
Notwithstanding the veneration of absinthe that pervaded this entire artistic epoch, one wonders how much the "doors of perception" (to borrow from Aldous L. Huxley's 1954 essay) were opened for these creative people. The artists were not constantly intoxicated, and indeed, there is good evidence to indicate that those works regarded as outstanding were usually created in lucid moments. On the other hand, the novel experiences of relative sizes, shapes and colors perceived under the influence of absinthe could have been recalled later and incorporated in a new font, palette or composition.

Oil of wormwood and alcohol were the standard ingredients of absinthe. The flavor and color of the drink were augmented with extracts of various plants: anise, fen-

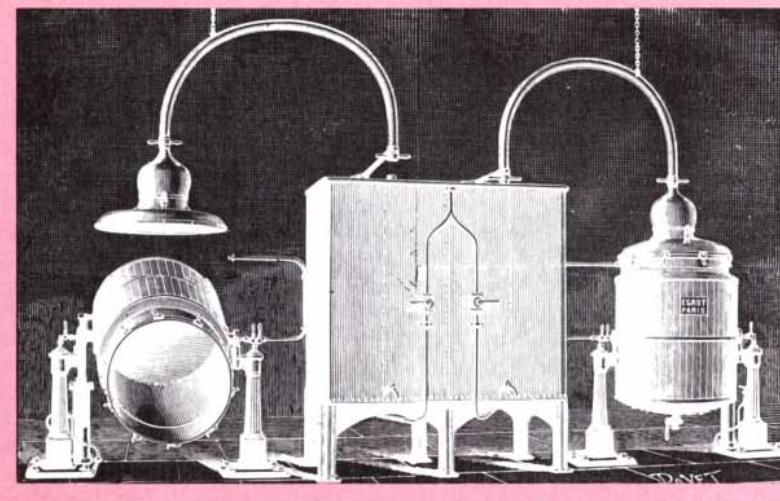
nel, hyssop, melissa (lemon balm) and, to a lesser extent, angelica, dittany of Crete, juniper, nutmeg, star anise and veronica—to name but a few. Specifications varied with the region and the maker. The general procedure for making absinthe involved steeping the mixture of herbs in a strong alcohol solution and then distilling the alcohol together with volatile constituents [*see box on this page*]. Absinthe was also made by adding individual essential oils to grain alcohol, a method more convenient for concoction on demand. George Saintsbury, a fin-de-siècle English literary critic and commentator on matters alcoholic, wrote that "almost every French chemist [pharmacist] in every small town had a liqueur of his own which was sovereign for digestion and other things."

The high ethanol content was not the special health hazard of absinthe, because it was diluted with water; the concentration of alcohol in diluted absinthe was certainly no greater than that in drinks containing brandy, whiskey, gin or rum. The main function of the alcohol concentration was to keep the oil constituents in solution. The *louche,* or turbidity, resulting from dilution was caused by the terpenes in absinthe, which came out of solution when the alcohol concentration was lowered and formed a colloidal suspension. These terpenes included thujone, fenchone, pinocamphone and citral [*see illustration on next page*]. Modern methods of analysis, employing gas chromatography coupled with mass spectrometry, have identified several additional terpenes and other chemicals in the essential oils that were incorporated into absinthe.

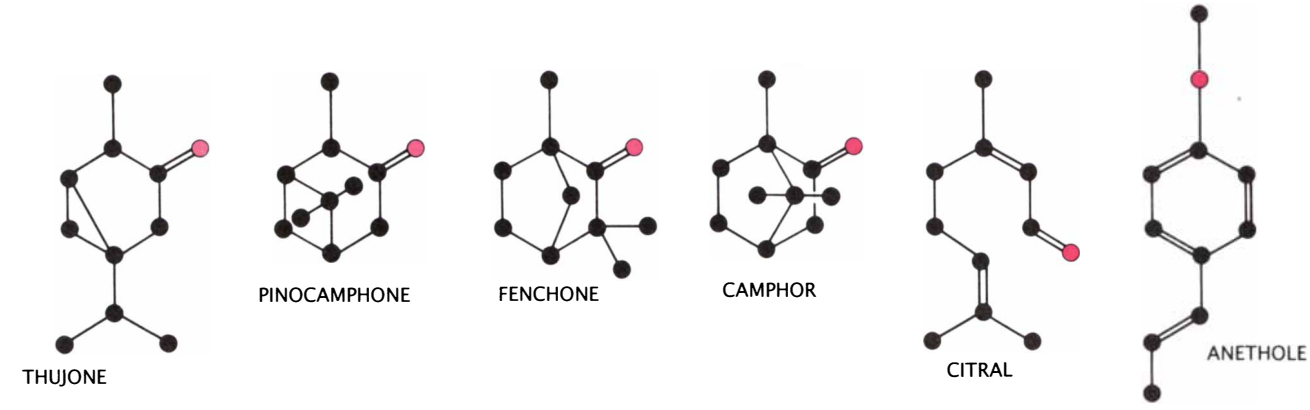
Some batches of absinthe contained questionable ingredients. Whereas the green coloration of properly prepared absinthe came from chlorophyll, there were reports of copper salts being added to inferior batches to improve the tint. On the bases of color and solubility the most likely adulterant was normal cupric acetate. Other reports pointed to occasional contamination with methanol and alcohols higher than ethanol, although the same could be said of other liqueurs. A report of antimony in some batches prompted a medical annotation in the *Lancet* in 1873 speculating, with better intentions than wisdom, that tartar emetic was also added in an attempt to make the drink less toxic. That salt, however, is only sparingly soluble in alcohol. A better candidate is antimony trichloride, which is poisonous. It is soluble in alcohol and produces a creamy precipitate on dilution with



DILUTION of the alcohol concentration of absinthe caused precipitation of a colloidal suspension of terpenes, of which thujone is one. The ritual of presentation involved pouring cold water over a sugar cube placed on a slotted spoon. (Sugar took the edge off the liqueur's bitterness.) One can get the same visual effect by diluting today's nontoxic substitute, pastis; the author did this demonstration with Pernod.



Large steam-heated stills such as the pair shown were used in the production of absinthe during the second half of the 19th century. An 1855 recipe from Pontarlier, France, gives the following instructions for making absinthe: Macerate 2.5 kilograms of dried wormwood, 5 kilograms of anise and 5 kilograms of fennel in 95 liters of 85 percent ethanol by volume. Let the mixture steep for at least 12 hours in the pot of a double boiler. Add 45 liters of water and apply heat; collect 95 liters of distillate. To 40 liters of the distillate, add 1 kilogram of Roman wormwood, 1 kilogram of hyssop and 500 grams of lemon balm, all of which have been dried and finely divided. Extract at a moderate temperature, then siphon off the liquor, filter and reunite it with the remaining 55 liters of distillate. Dilute with water to produce approximately 100 liters of absinthe with a final alcohol concentration of 74 percent by volume.



COMPOUNDS IN ABSINTHE include several terpenes: thujone (from wormwood), pinocamphone (from hyssop) and fenchone (from fennel). These compounds are ketones and structural isomers of camphor ($C_{10}H_{16}O$). The aliphatic aldehyde citral ($C_{10}H_{16}O$), which is also a terpene, comes from melissa. Fennel and anise also contribute an aromatic ether called anethole ($C_{10}H_{12}O$). The carbon atoms in this illustration are black and the oxygen atoms, red; hydrogen atoms are not shown.

water and most likely was added to get a better *louche* effect.

Such malpractices certainly added insult to injury for the absinthe consumer, but even the "best" of absinthe was toxic enough. Several of the early herbals warn against excess, but it is Johan Lindestolophe's *De Venenis* ("On Poisons"), published in 1708, that states clearly for the first time that continued use of *Artemisia absinthium* will lead to "great injury of the nervous system." The author and his commentator Christianus Stenzelius attested to the narcotic and debilitating effects of the herb from personal experience.

In 1859 Auguste Motet completed his thesis for the medical degree, "On Alcoholism and the Poisonous Effects Produced in Man by the Liqueur Absinthe." The title was prophetic, but given the vehicle the report probably did not reach the audience it deserved. In 1864, however, one of the leading journals of the day published a short note in which Louis Marcé of the Bicêtre, the famous hospital in Paris, described experiments with dogs and rabbits given essence of absinthe. The treated animals had suffered convulsions, involuntary evacuations, abnormal respiration and foaming at the mouth. Marcé juxtaposed these symptoms with those experienced by absinthe drinkers. He clearly understood the "double action" of absinthe intoxication: the separate effects caused by alcohol and thujone.

Marcé's student and collaborator, Valentin Magnan, carried those studies forward at the Saint Anne Asylum, focusing on the different effects engendered by absinthe as opposed to alcohol alone. Magnan and his colleagues observed that absinthe could cause hallucinations (both auditory and visual) in human beings, and they also induced them in experimental animals. For example, dogs given absinthe would posture toward a blank wall as if confronting imaginary foes. A single dose, albeit a fairly large one, caused convulsions mimicking those of epilepsy. Oil of wormwood elicited all the hallmarks of absinthism, and controlled experiments excused other essences within absinthe.

The 1865 edition of *Dictionnaire de Médecine,* by M. P. Émile Littré and Charles P. Robin, listed absinthism as a variety of alcoholism but emphasized that the special neurological effects were attributable to something other than alcohol. In 1868 Robert Amory, one of Magnan's students, gave a summary on absinthism to the *Boston Medical and Surgical Journal,* which is now the *New England Journal of Medicine.* In 1874 Magnan reviewed his papers on the subject in the *Lancet.*

The scientific warnings eventually reached the popular press, but they were matched by denials from those with an economic interest in the liqueur. Men and women caught up in the industrial revolution enjoyed the release absinthe provided and strove to convince themselves that the risks were small. Consumer reaction at the turn of the century ranged from mild restraint in drinking the spirit to complete disdain for the medical allegations that had been leveled against it.

In the period from 1875 to 1913, the annual consumption of absinthe per French inhabitant increased 15-fold. France imbibed about 10.5 million gallons of absinthe in 1913. There were regional differences: in and around Arles, for example, the rate was four times the national average. Statistics showed significant positive correlations between per capita absinthe consumption by region and the incidences of neurological disorders, stillbirths and rejections of army conscripts because of psychoses. Heinous crimes were blamed on absinthe intoxication.

Several attempts at reducing consumption of absinthe by increasing taxation were to no avail, and so in 1912 the French government demanded that the concentrations of both alcohol and essential oils be lowered. Consumers merely modified the ratio of water to absinthe. The unregulated sale of essences and powders of wormwood under trade names and the availability of cheap (and possibly contaminated) alcohol routed the legislative exercise. A prohibition on both the sale and manufacture of absinthe in France was formalized in 1915, but there was some vacillation, and the ban was not reasonably enforced until some years later. Belgium, Switzerland, the U.S. and Italy took similar actions between 1905 and 1913.

In 1901 Raoul Ponchon, bon vivant and commentator on all elements of Parisian life, wrote a poem called *Absinthe and the Guinea Pig,* after a report that the deputy chief of the municipal laboratory had injected an animal with 10 milliliters of absinthe to illustrate the mortal toxicity of the poet's favorite drink. The "scientific" demonstration had been intended more for dramatic effect than for accurate simulation of the human habit, but the poet picked up on the concept of dosage and remarked that he would have to drink a liter of the beverage at a single sitting in order to match the guinea pig's "binge."

Although Ponchon's skepticism became less organized as the poem progressed, his initial point is well taken. It is still not clear how much absinthe was too much. The human dose-re-

sponse relation is difficult to assess from available data. The binge drinker experienced hallucinations from acute intoxication; the chronic imbiber suffered some irreversible brain damage to an extent dependent on the amount of absinthe and the frequency of consumption as well as the age, nutritional status and general health of the imbiber. Upset stomachs were common in habitual drinkers, particularly those whose diets were less than adequate.

The correct chemical structure of thujone was published in 1900 by the German chemist Friedrich W. Semmler, and by 1916 European and American scientists had documented its pharmacodynamics. The compound causes marked excitement of the autonomic nervous system, followed by unconsciousness and convulsions. The involuntary and violent muscular contractions are at first clonic (rapid and repeating with intervening relaxation) and then tonic (continuous and unremitting). The effects of thujone are practically identical with those of camphor. Camphor- and thujone-induced convulsions were studied as a model for epilepsy; a number of research papers describing these studies appeared in the neurology and psychiatry journals during the 1920's and 1930's.

Camphor was subsequently employed by László J. von Meduna and his colleagues at the National Hospital for Nervous and Mental Diseases in Budapest in convulsive therapy for certain cases of schizophrenia. Early difficulties in dosage regulation and the side effects of intramuscular camphor injections were avoided by substituting first intravenous pentylenetetrazole and then inhalation of hexaflurodiethyl ether. Although electroconvulsive therapy has replaced these chemical approaches, it is still remarkable to note that the beneficial effect of such therapies is brought about by the convulsion itself rather than by the compound administered or the flow of current. In this area thujone and camphor played positive roles in the evolution of an important medical practice.

The concentration of ethanol in spirits is customarily expressed as a percentage or a proof value. A 70 percent solution by volume consists of 70 volumes of pure ethanol brought to 100 volumes of solution by the addition of water; because of the interaction between ethanol molecules and water molecules, 33.4 volumes of water, rather than 30, must be added. The most practical means of determining the alcohol concentrations of commercial spirits is the alcoholometer shown here, a device perfected in 1824 by Joseph-Louis Gay-Lussac. The meter measures the specific gravity of a distillate; it is floated in a sample at a prescribed temperature, and the depth to which it sinks indicates the percentage of alcohol in the solution. Percentage or proof value can be read directly from the scales inscribed on its stem. When absinthe was popular, alcohol concentration was expressed in the old French system of degrees, which were numerically equivalent to percentages by volume. The British coined the term "proof"; their proof spirit, which corresponded to 57.27 percent ethanol by volume, was operationally defined as one having the maximum concentration of water that would still allow ignition after admixture with gunpowder. In the U.S. the term was modified; the proof value is exactly double the percent by volume, so that a liquor that is 50 percent ethanol by volume is defined as 100 proof.

From the first to the 18th century, Chinese scholars extolled the virtues of the *qing-hao* plant (*Artemisia annua*) in the treatment of malaria. The efficacy of decoctions of this species was reconfirmed in 1971, and the active principle was identified the next year: it is an unusual sesquiterpene lactone peroxide named qinghaosu. Derivatives of the compound that are more effective have since been synthesized; their utility against otherwise resistant strains of the malarial parasite is an exciting development. Malaria was common around the Mediterranean region in the 19th century, and for a time I wondered whether the French troops of the 1840's might have arrived at a preventive medicine when they added wormwood to their daily wine. But it turns out that *Artemisia absinthium* does not contain enough qinghaosu to make it a significant source of the compound.

Artemisia species have been praised as sources of insect repellants, anthelmintics and antimalarials. Thujone and its chemical cousin camphor played roles in basic research on epilepsy and convulsive therapy that were judged positive and constructive. Absinthe tippling, on the other hand, was judged to be negative and destructive, and in retrospect, the interdiction was tardy but surely justified. Opinions to the contrary have been expressed in the past 20 years, but they seem to be based on romantic and wishful thinking. After the ban on absinthe a substitute containing no wormwood and additional anise was offered on the Continent; two of the proprietary names are Ricard and Pernod.

FURTHER READING

ON THE COMPARATIVE ACTION OF ALCOHOL AND ABSINTHE. V. Magnan in the *Lancet*, Vol. 2 for 1874, No. 2664, pages 410–412; September 19, 1874.

THE DICYCLIC TERPENES, SESQUITERPENES AND THEIR DERIVATIVES, VOL. 2. John Lionel Simonsen in *The Terpenes*. Cambridge University Press, 1949.

AN ILLUSTRATED HISTORY OF THE HERBALS. Frank J. Anderson. Columbia University Press, 1977.

ABSINTHE: HISTORY IN A BOTTLE. Barnaby Conrad III. Chronicle Books, 1988.

VINCENT VAN GOGH AND THE THUJONE CONNECTION. Wilfred Niels Arnold in the *Journal of the American Medical Association*, Vol. 260, No. 20, pages 3042–3044; November 25, 1988.

THE AMATEUR SCIENTIST

The mechanics of rock climbing, or surviving the ultimate physics exam

by Jearl Walker

Rock climbing might be said to be the ultimate physics exam, because failure can lead to death; even partial credit can easily result in grave injury. Certainly the sport demands not only strength and agility but also an intuitive understanding of physics. The essential principles are handed down as rules from instructors to students, and most of them are easy enough to follow. The significance and applicability of a few of them are not always obvious, however; a novice climber can be tempted to disobey one or another of the rules.

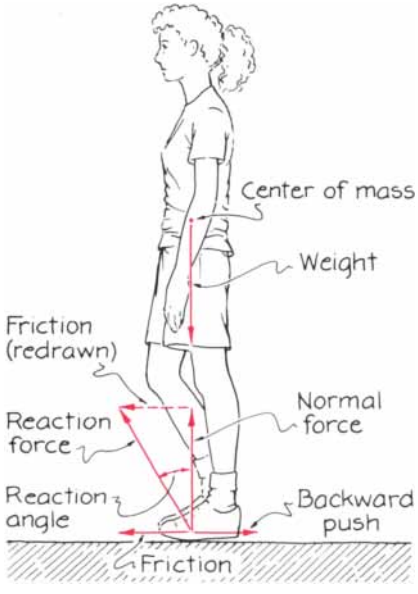
Imagine, for example, that you are climbing without benefit of equipment other than a safety belay line handled by a companion above you. You face a tilted slab of rock that is met at an obtuse angle by another slab of similar tilt, the juncture forming a ridge something like an edge of a pyramid. (The juncture is not sharp, though; it is worn as smooth as the faces of the slabs.) Where is the easiest climb: directly up one of the slabs or along the ridge?

Again, suppose that the grade on a slab is steep but would nevertheless allow you to stand and manage to walk up the rock. Is it safer to lean forward and use your hands? What if you are on a sloped ledge and find a similarly sloped ledge at about waist level: is it safer to lean forward and partially support yourself with your hands on the higher ledge?

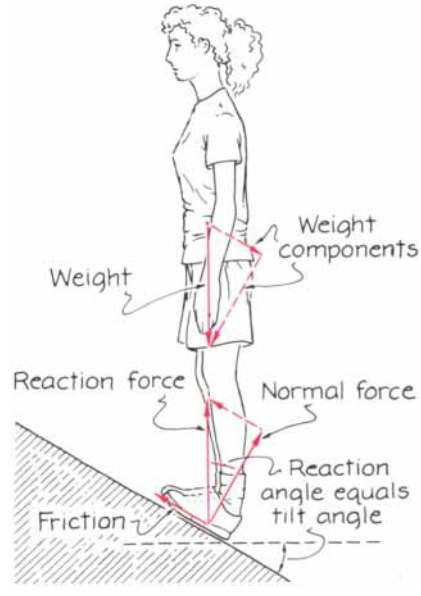
Consider a third situation. A fairly wide vertical crack might allow you to "chimney": to climb into the crack, push your back against one side and press your feet against the opposite side, bending your knees to make the fit. To move upward you might essentially walk up one side of the crack while shoving your back up along the other side. (You gain additional, but minor, support if you press your hands against the wall at your back or, if the crack is narrow enough, against the opposite wall.) Or you might place one foot on each wall and press down while raising your body. Because you must exert constant pressure against the walls, a long chimney is tiring. If you stop to rest, is there an optimal location for your feet, one that would minimize the necessary push on the rock? When you climb, how exactly should you move your feet and back to decrease the chance of slipping? Suppose that as you climb you find the rock becoming slick. Should you place your feet higher or lower with respect to where your back pushes on the rock?

If, to cite a final example, the crack is quite narrow and the rock on one side of it juts out more than the other, you might try a "lie-back." The technique requires that you be on the side of the crack opposite the section that juts out. You insert your fingers into the crack and pull against its side. Then you place your feet on the projecting rock on the opposite side and press hard. As you pull with your arms and push with your legs, the tension you create stabilizes you, but the effort is draining. How should you position your arms and legs to lessen the fatigue, and how should you move them to shift yourself upward safely?

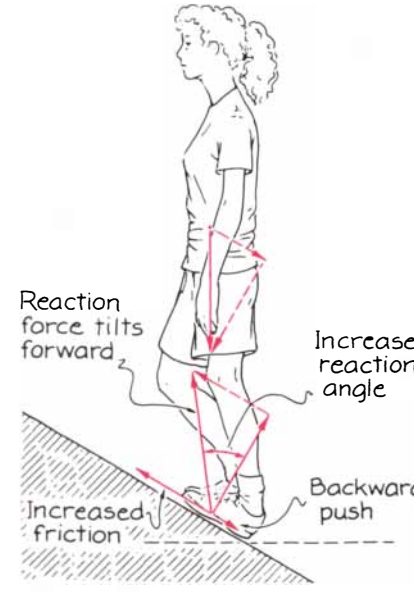
Some of the principles involved in these varied techniques of climbing were examined in 1976 by R. R. Hudson of the Darlington College of Technology and W. Johnson of the University of Manchester Institute of Science



Forces involved in a backward push



Forces arising from a tilted surface



A backward push on a tilted surface

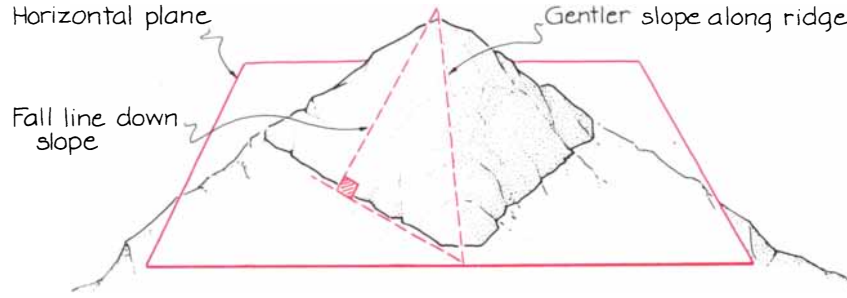
and Technology. The techniques rely on the friction between the climber and the rock. Friction, of course, is the force that opposes the sliding of one surface over another, such as a climber's shoe over rock. A simplified explanation of friction involves the small-scale roughness of the surfaces. Even if two surfaces, such as flat rock and the sole of a shoe, appear to be smooth, they are likely to be covered with tiny hills and valleys. When a force tries to slide the surfaces over each other, the hills on one surface catch on the hills and in the valleys of the opposite surface. The collective resistance to the sliding motion is the friction.

Before I get to the tricky cases of rock climbing explored by Hudson and Johnson, I shall examine the forces involved in some common situations of standing and walking. When you stand upright on a floor, two forces operate on you: your weight pulls down, and the floor pushes up just as much. The forces can be represented by vectors that indicate size and direction [*see illustration at left on opposite page*]. The weight vector is assigned to the center of mass, which is approximately behind the navel when you are upright. The force from the floor is called the normal force, "normal" indicating that the force is perpendicular to the support surface—in this case the floor.

If you begin to step forward with one foot, your leg muscles push backward on the other foot, which experiences friction that must exactly counter the push if it is not to slip. The friction is represented by a vector parallel to the floor. If you push harder and the trailing foot still does not move, the friction must then be larger and, again, an exact match to the push. If you push too hard, so that the push exceeds some upper limit on the friction, the hills will yield and your trailing foot will slip. The upper limit on the friction is determined by the product of the normal force and the "coefficient of friction," a measure of the roughness of the sole and the floor.

Because the normal force equals your weight, the upper limit on the friction depends primarily on the coefficient of friction. When the coefficient is large, as it is on dry concrete, the upper limit is also large; you may be able to push quite hard without slipping. When the coefficient is small, as it is on slippery ice, even a small push makes the trailing foot slide.

An equivalent way to examine the role of friction is to redraw the friction vector so that it and the normal force form the legs of a right triangle. The procedure does not change the value of the friction, because the essence of the vector—its length and orientation—is retained. The hypotenuse of the triangle, also a vector, is said to be the sum of the normal force and the friction; it is called the reaction force.

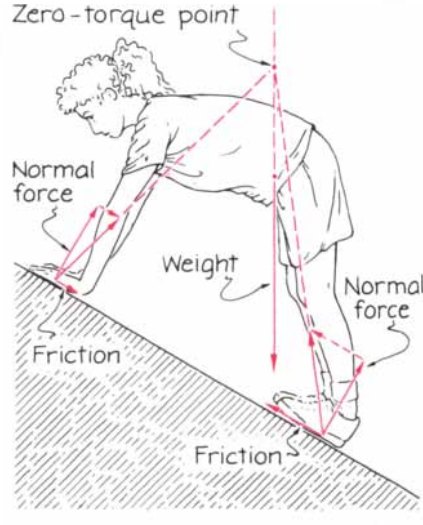


Why the ridge is less steep

The angle between the reaction force and the normal force—what I shall call the reaction angle—is important. As you push harder and increase the friction, the reaction angle increases. When the angle reaches a certain critical value, the friction is at its maximum value and the trailing foot is on the verge of slipping. The critical case arises when the tangent of the angle is equal to the coefficient of friction. In some examples of rock climbing the reaction angle is easier to study than the friction on the climber, and so I shall be considering it further.

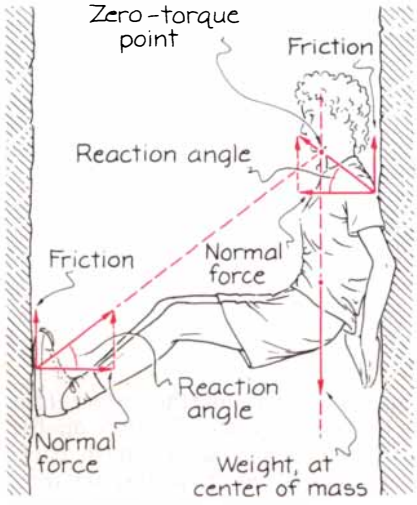
Now consider the forces involved when you stand upright on a ramp [*see middle illustration on opposite page*]. Your weight still pulls you toward the support surface and leads to a normal force that is perpendicular to the surface, but now your weight also pulls you down along the surface. That new feature creates friction on the feet even if you do not push backward. An easy way to picture the forces is to separate the weight vector into its components. One component is perpendicular to the ramp and is matched by the normal force. The other component points down the ramp, and if you are stable it is matched by the friction. Here again the friction can be redrawn so that it, the normal force and the reaction force form a right triangle. In this example the reaction angle happens to be the same as the tilt of the ramp.

Imagine what happens if the tilt (and so also the reaction angle) increases and you remain upright. The weight component directed down the ramp increases and so, automatically, does the friction that counters it. The weight component perpendicular to the ramp decreases, as does the normal force that opposes it. When the reaction angle reaches some critical value, the friction reaches its upper limit and you are on the verge of slipping. That critical value sets a limit on the pitch you can withstand without slipping.



Forces when a climber leans onto a slope

Leaning onto a second ledge



Forces in a chimney climb

If the pitch is moderate and you push backward on one foot in order to step up the ramp with the other foot, the push is directed down the ramp and requires that the friction on the trailing foot increase. This tilts the reaction force forward from the vertical and increases the reaction angle [*see illustration at right on page 118*]. The stability of the foot can be described in terms of either the friction or the reaction angle. The foot is stable if the friction does not exceed its upper limit or, equivalently, if the reaction angle does not exceed its critical value. If you carelessly push too hard, your foot can slip even if the pitch is only moderate.

Armed with these examples, I now return to the specific questions about rock climbing. First was the question of whether to scale a tilted slab directly or along its juncture with another, similarly tilted slab. On either route you always need to keep the reaction angle below its critical value, which is set by the coefficient of friction between your shoes and the rock. The danger of slipping is greatest when you step up the slope and push backward on the trailing foot. To be safe you should choose the route with the least pitch, even when you might be able to stand stably on another route with a steeper grade. The least pitch is along the ridgelike juncture. One way to see this is to run a line straight down a slab so that it is the leg of a right triangle, the hypotenuse of which is the ridgelike juncture [*see top illustration on preceding page*]. The leg, being shorter than the hypotenuse, must have the steeper pitch.

When you can stand upright on a steep slope, should you bend over to use your hands? Most rock-climbing instructors teach that you should always stand upright if possible and that bending over greatly increases the chance your feet will slip. Are the instructors right? Friction must still counter the weight component that is directed down the slope, but now there is friction on both hands and feet. Would that not be safer than standing upright and depending only on the friction on the feet?

Hudson and Johnson sided with the instructors for several reasons. Your hands will encounter some friction, but the coefficient of friction between skin and rock is small, and so the upper limit on that friction is also small. If your hands slip the only force that can stop a full slide down the slope is the friction on your feet. Although there is enough friction to hold you if you stand upright, when you lean over there very well may not be enough.

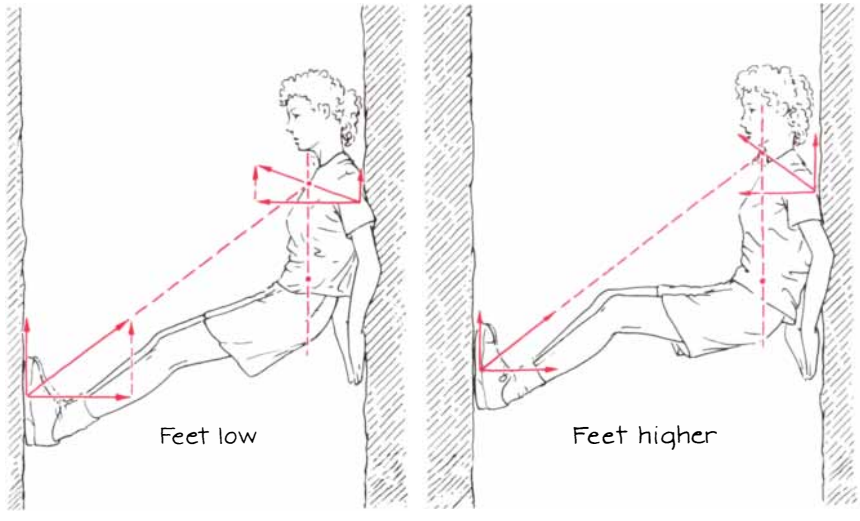
To explain the result I must review the requirements of stability and introduce a new one. So far I have satisfied three requirements. The forces up and down the slope must balance; the forces perpendicular to the slope must balance; and friction must not exceed its upper limit or, equivalently, a reaction angle must not exceed its critical value. The new requirement is that the torques on a climber must balance. (A torque is the tendency of a force to make an object rotate about some pivot point.) For a climber to be stable the torques from the weight and the reaction forces must balance for every possible choice of pivot.

The requirement is easily simplified. For example, suppose the climber has leaned forward to put his or her hands on a uniform slope [*see bottom left illustration on preceding page*]. The torques are guaranteed to balance if extensions of the weight vector and the reaction forces on the hands and feet pass through a common point. To picture this condition, start with the weight vector. It is vertical and passes through the climber's center of mass, which lies somewhere near the navel (the actual location depends on the configuration of the body). Mentally draw a vertical line through the center of mass.

Now consider the reaction forces. If the climber is stable extensions of those forces must meet at some point on the vertical line—a point I shall call the zero-torque point, because the torque from each force is zero if the pivot is chosen to be there. If the forces do meet at a zero-torque point, then the climber is stable against rotation for any other choice of pivot. I do not mean to imply that the climber consciously runs through a calculation to see if this requirement is met. Rather, I mean that if it is not met, the climber feels the instability and must alter the distribution of weight, the extent of leaning or the tension in muscles to reattain stability.

Is the leaning orientation shown in the illustration safer than an upright stance? Notice that for stability against torques the reaction force on the feet must tilt forward if it is to pass through a zero-torque point. Hudson and Johnson argued that the reorientation of the reaction force increases the reaction angle at the feet and that, if the slope is steep, the angle may exceed its critical value. And so to lean forward is folly, even if there is friction on the hands.

I thought about the matter further.



How forces in a chimney depend on placement of the feet

Suppose the hands are positioned a certain distance up the slope from the feet. The sum of the friction on hands and feet must counter the weight component down the slope, but exactly how much friction is on the hands and how much is on the feet? Without knowing details about the muscular forces inside the body the question cannot be answered. There are four different variables involved in the example—two normal forces and two frictional forces—but we have only three equations in which forces and torques are balanced. Three equations are insufficient to determine four variables. If you plop your hands down on the slope, there is no practical way to predict how much friction will be on your hands or feet.

I identified still another danger in leaning forward onto a slope. Suppose the hands are placed well up the slope. If the reaction force on them is to pass through a zero-torque point, the friction on them must be directed down the slope. That in turn requires that the friction on the feet be even larger than when you stand upright, for now it must balance the weight component down the slope plus the friction on the hands. The situation seems no less dangerous if the climber leans well forward to put the hands on a second sloped ledge at about waist level. In either case the danger lessens if the hands are kept low and the lean small, but as a rule the climber should obey the instructor and avoid leaning well forward onto the rock.

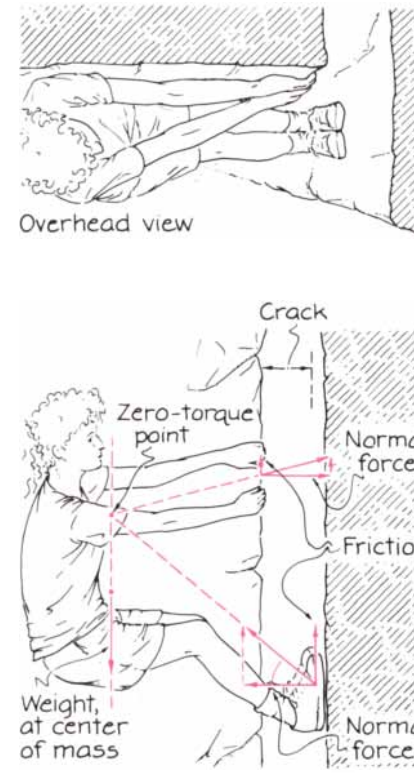
I next consider the chimney climb [*see top illustration on opposite page*], which as a youth I did routinely in various caves of West Texas. For stability in this maneuver the sum of the friction on the feet and the shoulders balances the climber's weight, and the normal forces on the feet and the shoulders match. (This time there are only three different variables, and the equations of balance can be solved.) The climber's exertion is reflected by the size of the normal force, and so during a rest the goal is to reduce that force as much as is safely possible. Hudson and Johnson stated that the normal force can be minimized if the feet are positioned at a certain distance below the shoulders and if both feet and shoulders are on the verge of sliding. As you might guess, the distance depends on the coefficients of friction at the feet and the shoulders. If either coefficient should decrease—because of wetness, say—the climber must decrease the vertical distance between feet and shoulders.

To check Hudson and Johnson's conclusion I went through the following mental exercise. First, how does the friction on the feet depend on where the feet are placed? Consider the climber's weight and the reaction forces on the feet and the shoulders [*see bottom illustration on opposite page*]. Again, stability requires that extensions of the forces pass through a zero-torque point. Imagine what happens to the reaction force on the feet if they are moved and the normal force is unchanged. If the feet move downward the friction on them must increase if the reaction force is still to extend up toward a zero-torque point. If the feet move upward the converse happens. Also, because the sum of the forces must balance the climber's weight, any change in the friction at the feet must be met by an opposite change in the friction at the shoulders.

Suppose the climber places the feet at some low but stable point on the chimney and relaxes the push against the rock until the feet are on the verge of slipping. The reaction angle at the feet is then at its critical value. If the climber were to move the feet downward, the friction on the feet would increase; to keep the reaction angle from exceeding its critical value, the normal force would have to be increased. And so moving the feet down is not a good idea.

What about moving the feet upward? The friction on the feet then decreases, and the reaction angle can be kept at its critical value while the normal force is safely decreased. Moving the feet up may allow the climber to relax somewhat, but because of changes in the forces at the shoulders, the feet cannot be lifted too high. As the feet are raised and the climber relaxes, the friction at the shoulders edges toward its upper limit. The normal force and the push by the climber can be minimized when the friction at the shoulders reaches the upper limit. If the feet are lifted any higher, the normal force and push must increase.

To move up a chimney a climber must push downward on one foot, but the push requires additional friction on the foot and additional normal force if the foot is not to slip. To be safe the climber should prepare for an upward move by first raising one foot and bracing it against the wall. If the lower foot happens to slip during the ascent, support by the higher foot can prevent a fall. The same technique can make a lie-back safer: when the climber must push downward on a foot to raise the body, the other foot should be poised higher in case it must take over if the lower foot slips.



A lie-back on a narrow crack

A lie-back is tiring primarily because of the great tension on the arms [*see illustration above*]. If the climber wants to rest and minimize the tension, should the feet be high or low? You can explore the question by balancing forces and requiring that the reaction forces and weight extend through a zero-torque point.

In their own determination of the best location for the feet, Hudson and Johnson made several simplications. They assumed that the friction on the hands is small enough to ignore and that the friction on the feet fully supports the climber's weight. They also assumed that the reaction force on the hands is horizontal. The zero-torque point is then at the intersection of a vertical line through the center of mass and a horizontal line through the hands. If the climber raises the feet and they are still to supply the same amount of friction, what must happen to the normal force to keep the reaction force at the feet aimed at the zero-torque point?

FURTHER READING

ELEMENTARY ROCK CLIMBING MECHANICS. R. R. Hudson and W. Johnson in *International Journal of Mechanical Engineering Education,* Vol. 4, No. 4, pages 357–367; 1976.

COMPUTER RECREATIONS

A potpourri of programmed prose and prosody

by A. K. Dewdney

"'Take care of the sense and the sounds will take care of themselves.'"
—The Duchess, in *Alice's Adventures in Wonderland,* by Lewis Carroll

As the renowned Oxonian master of nonsense observed, semantics takes precedence over syntax in creative writing. Good literature is shaped by the meaning contained in a writer's ideas. Computers are not yet capable of ideas and so cannot take care of a composition's sense. Yet, as a number of contemporary programs show, computers can certainly take care of the sounds. But is it art? That is for the reader to decide.

Consider the writings of MARK V. SHANEY, a computer program created by Bruce Ellis (now at the University of Sydney) that is based on an idea of Don P. Mitchell of the AT&T Bell Laboratories. MARK V. SHANEY is not exactly a self-starter. The program must first read and reflect on someone else's work. It then produces a rambling and somewhat confused commentary on the work. As an example, I quote MARK V. SHANEY's reflections on reading an elementary chemistry textbook.

"Wash the blackboard. Watch it dry. The water goes into the air. When water goes into the air it evaporates. Tie a damp cloth to one end of a solid or liquid. Look around. What are the solid things? What are the only ones that take part in the clouds themselves. As clouds move air tosses the tiny droplets of water. Sometimes the water which leaves the body in the leaves of green plants. It is quite porous, almost like a sponge. But the walls of the hollow tubes and cells are very strong. Chemical changes take place when something burns . . . ."

The commentary is hardly different when MARK V. SHANEY digests a book on elementary mathematics.

"Why do we count things in groups of five. When people learned how to count many things, they matched them against their fingers. First they counted out enough things to match the fingers of both hands. Then they put these things aside in one quart. A giant-size bottle that will hold four quarts is a three-digit number . . . ."

From here on, the program's output becomes increasingly confused, as readers may see for themselves in the illustration on this page.

Although sense is conspicuously absent from MARK V. SHANEY's writings, the sounds are certainly there. The overall impression is not unlike what remains in the brain of an inattentive student after a late-night study session. Indeed, after reading the output of MARK V. SHANEY, I find ordinary writing almost equally strange and incomprehensible!

How does MARK V. SHANEY produce its remarkable works? The answer is rather simple. The program's name, a weak pun on "Markov chain," provides a clue. In abstract terms, a Markov chain is a sequence of symbols generated according to a table of probabilities. In the version relevant to MARK V. SHANEY's operation, each row of the table corresponds to a pair of symbols. The entries in each row consist of individual symbols, each paired with an associated probability. A sequence of symbols is generated by an algorithm that begins with a "chain" of two symbols and thereafter cycles through four simple steps.

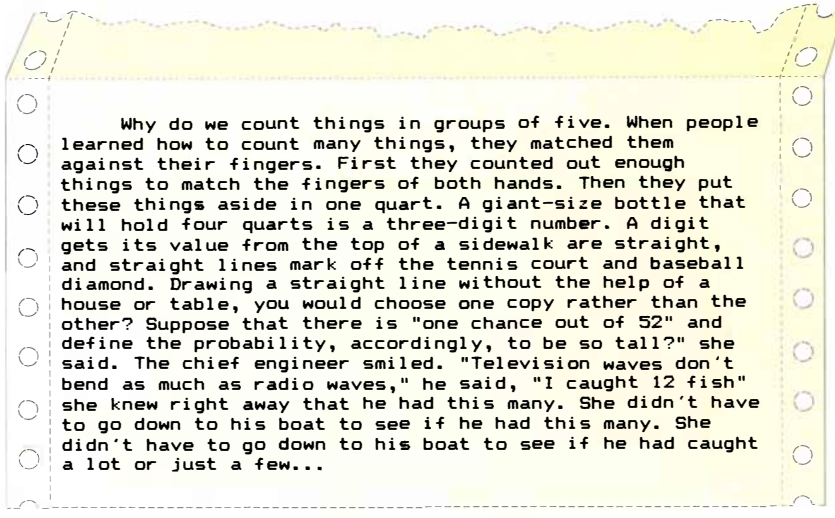
1. Find the last two symbols in the current chain.
2. Go to the row of the table corresponding to the symbol pair.
3. Select a symbol from the row according to its probability.
4. Add the selected symbol to the end of the chain.

For example, the first few entries of a Markov-chain table for the symbols *A, B, C* and *D* might look like this:

AB: *B*(.1) *C*(.1) *D*(.8)
AC: *A*(.1) *B*(.2) *C*(.1) *D*(.6)
AD: *B*(.4) *D*(.6)
BA: *B*(.3) *C*(.4) *D*(.3)
BB: *A*(.5) *C*(.5)

.
.
.

Given the symbol pair *AB* as the initial chain, the algorithm would have a 10 percent chance of selecting *B,* a 10 percent chance of selecting *C* and an 80 percent chance of selecting *D* as the next symbol. How does the algorithm choose a symbol according to the probabilities? It divides the interval between 0 and 1 into numerical



Why do we count things in groups of five. When people learned how to count many things, they matched them against their fingers. First they counted out enough things to match the fingers of both hands. Then they put these things aside in one quart. A giant-size bottle that will hold four quarts is a three-digit number. A digit gets its value from the top of a sidewalk are straight, and straight lines mark off the tennis court and baseball diamond. Drawing a straight line without the help of a house or table, you would choose one copy rather than the other? Suppose that there is "one chance out of 52" and define the probability, accordingly, to be so tall?" she said. The chief engineer smiled. "Television waves don't bend as much as radio waves," he said, "I caught 12 fish" she knew right away that he had this many. She didn't have to go down to his boat to see if he had this many. She didn't have to go down to his boat to see if he had caught a lot or just a few...

MARK V. SHANEY's mathematical commentary

segments whose lengths correspond to the symbol probabilities. It then chooses a random number between 0 and 1 and determines in which segment the number has fallen.

For row *AB* in the above table, the segments corresponding to the respective probabilities for *B, C* and *D* range between 0 and .1, between .1 and .2, and between .2 and 1. Suppose, then, that the computer's random-number generator yields .0172. Because that number lies in the first segment, *B* would be selected as the next symbol in the chain. The chain would then consist of the symbols *ABB*. The algorithm would next consult row *BB* in order to select a fourth symbol for the chain. Here again, a random number is generated. If it is less than or equal to .5, *A* is selected; otherwise the algorithm selects *C*. Because of its dependence on chance, if the algorithm were restarted with the same initial symbol pair, it might well produce an entirely different chain.

Such an algorithm was actually applied in the 1940's by Claude E. Shannon of Bell Laboratories to analyze the information content of human language. He constructed the algorithm's probability tables by scanning ordinary text and counting how many times each individual character followed each pair of characters (including blanks). Once the character frequencies for a given text were known, they could easily be changed into probabilities. The Markov chains of characters generated in this manner had statistical properties that resembled the source text, although they rarely contained valid words. How then does MARK V. SHANEY apply Markov chains to produce understandable English words?

The trick is to apply Shannon's algorithm for Markov chains but with entire words instead of characters as the concatenated symbols. As MARK V. SHANEY scans a text, it builds a frequency table for all words that follow all the word pairs in the text. The program then proceeds to babble probabilistically on the basis of the word frequencies.

A key feature of the program is that it regards any punctuation adjacent to a word as part of the word. That feature enables it to form sentences having a beginning and an end. Approximately half of them are even grammatical. I shudder to think what the program might produce after scanning this article!

Indeed, others have already shuddered at MARK V. SHANEY's reflections, some with rage and others with laughter. Some years ago Ellis decided to go on-line with his creation. The victims of the program's analyses were the innocent computer users who subscribed to an electronic bulletin board called net.singles. The bulletin board provides a place for male and female scientists, engineers, programmers and graduate students from all over the country to post their thoughts on subjects as diverse as dating, makeup and personal relationships. Why not have MARK V. SHANEY read the postings and respond with its own "thoughts" on those subjects? Here are some of MARK V. SHANEY's more thoughtful comments.

"When I meet someone on a professional basis, I want them to shave their arms. While at a conference a few weeks back, I spent an interesting evening with a grain of salt. I wouldnt dare take them seriously! This brings me back to the brash people who dare others to do so or not. I love a good flame argument, probably more than anyone . . . .

"I am going to introduce a new topic: does anyone have any suggestions? Anybody else have any comments experience on or about mixed race couples, married or otherwise, discrimination forwards or reverse, and eye shadow? This is probably the origin of makeup, though it is worth reading, let alone judge another person for reading it or not? Ye gods!"

The opinions of the new net.singles correspondent drew mixed reviews. Serious users of the bulletin board's services sensed satire. Outraged, they urged that someone "pull the plug" on MARK V. SHANEY's monstrous rantings. Others inquired almost admiringly whether the program was a secret artificial intelligence project that was being tested in a human conversational environment. A few may even have thought that MARK V. SHANEY was a real person, a tortured schizophrenic desperately seeking a like-minded companion.

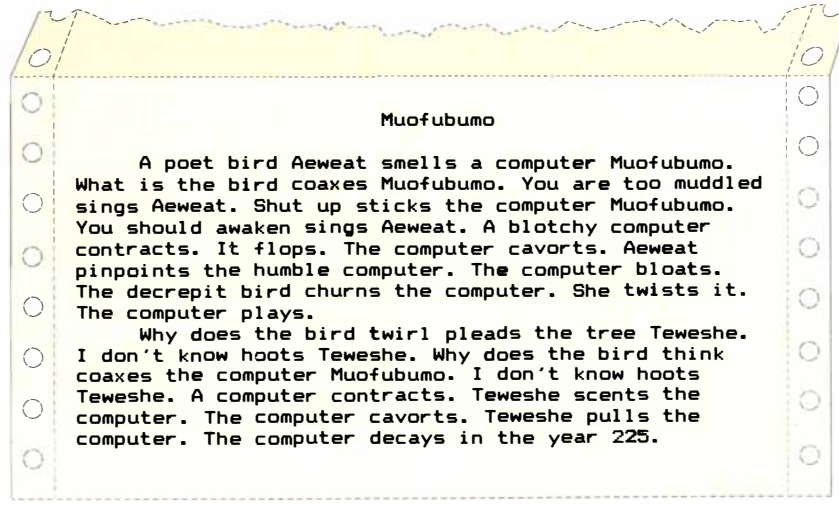
If the purpose of computer prose is to fool people into thinking that it was written by a sane person, MARK V. SHANEY probably falls short. MELL, the brainchild of Bonnie V. Firner of Piscataway, N.J., probably comes a little closer to that goal. MELL writes weird science-fantasy stories with a peculiar meditative quality.

"The warrior scowls in the drought pulses. He loves himself in the drought. He molds himself in the drought. He glowers at the warrior Dugaki in the drought pulses. He calls her in the drought. He snarls at her in the drought. He sits beside her beside awareness. He seizes her.

"Oh I am mighty says Oban. He smothers her in the drought. He smashes her of the turquoise . . . .

"Oban kills Dugaki. He has it of turquoise. He glares at it in the drought pulses. He calls it in the drought. He snarls at it in the drought. He sits with it beside greed. He seizes it. He smothers it in the drought. He smashes it of the turquoise. He needs it beneath the thunderbolt. His body burns beneath the thunderbolt . . . . "

MELL consists of some 1,500 lines of BASIC code; in contrast to MARK V. SHANEY, Firner's program is complicated even at the conceptual level. The program's main loop generates two characters in terms of 16 randomly generated "descriptors" whose values define qualities such as size, niceness, occupation, age and health, smell, commitment (from indifferent to fanatical) and even magical power.

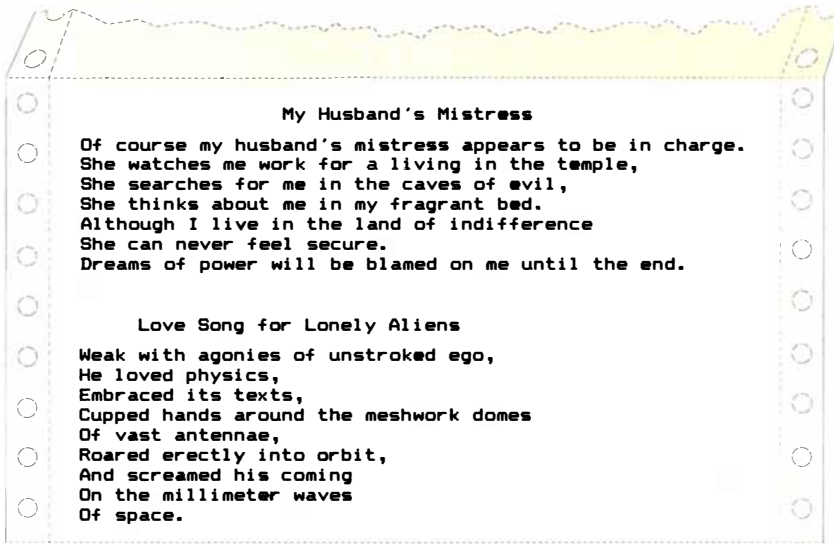


Muofubumo

A poet bird Aeweat smells a computer Muofubumo. What is the bird coaxes Muofubumo. You are too muddled sings Aeweat. Shut up sticks the computer Muofubumo. You should awaken sings Aeweat. A blotchy computer contracts. It flops. The computer cavorts. Aeweat pinpoints the humble computer. The computer bloats. The decrepit bird churns the computer. She twists it. The computer plays.

Why does the bird twirl pleads the tree Teweshe. I don't know hoots Teweshe. Why does the bird think coaxes the computer Muofubumo. I don't know hoots Teweshe. A computer contracts. Teweshe scents the computer. The computer cavorts. Teweshe pulls the computer. The computer decays in the year 225.

Firner's MELL tells a story



My Husband's Mistress

Of course my husband's mistress appears to be in charge.
She watches me work for a living in the temple,
She searches for me in the caves of evil,
She thinks about me in my fragrant bed.
Although I live in the land of indifference
She can never feel secure.
Dreams of power will be blamed on me until the end.

Love Song for Lonely Aliens

Weak with agonies of unstroked ego,
He loved physics,
Embraced its texts,
Cupped hands around the meshwork domes
Of vast antennae,
Roared erectly into orbit,
And screamed his coming
On the millimeter waves
Of space.

POETRY GENERATOR* (top) *and THUNDER THOUGHT* (bottom) *challenge the literati

Having chosen the names and traits of the story's characters, the main loop then determines what motivates a character by examining the values of its various descriptors. If one of the descriptors has a low value, MELL bases the character's interaction with the other characters on that fact. If no character has a low descriptor value, MELL will decide on the nature of a character's interaction based on his or her occupation. For each cycle of the main loop, the program then generates several sentences that describe the characters by fitting names for their qualities into predetermined grammatical slots. The sentences thus generated also include what Firner calls "background" words.

In the above story, for example, the background word "drought" sets the scene. Consequently that word creeps into many of the sentences. After generating a paragraph that describes an act by one of the characters in this way, the sentence-generating process starts all over again. Between iterations the program can change a quality of one of the characters. This helps to keep the story (as much as there is of one, anyway) from stagnating.

Is poetry any easier for a computer to generate than prose? One doubts it, because the meaning contained in prosody tends to be far more dense than in prose. Of the three versifying programs I shall discuss, only POETRY GENERATOR, which was written by Rosemary West of Mission Hills, Calif., is fully automated; the other two require human intervention to finish the product. West describes her program as follows:

"My approach ... was to supply a large vocabulary of words and phrases that would be selected at random and combined according to a set of grammatical rules. For example, consider the following poem: 'The tree dips its bare fingers / into the black ice-pond / just as three gray geese / slide down a nearby snowbank.'

"Each line of the poem can be broken down into several parts .... 'The tree' is a noun phrase; 'dips' is a verb; 'its bare fingers' is the object of that verb. Having categorized the parts, I can then come up with between 100 and 400 possible substitutions for each part, one of which is randomly selected by the computer. For example, using the same verse structure, I might get: 'A woman hides five gray kittens / under the old jalopy / at the moment when the sad clowns / enter your museum of pain.' "

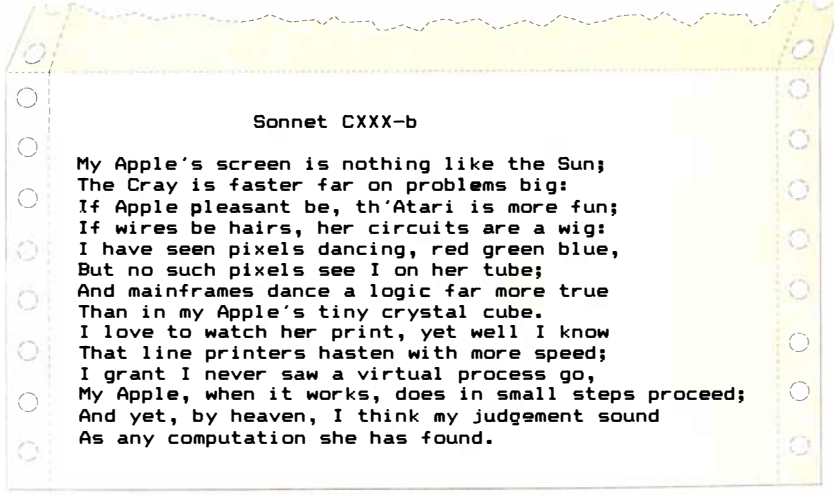
The poetic structures on which POETRY GENERATOR hangs its words may vary considerably, lending variety to the syntax and to what seems to be the meaning of the poem. West has based some of POETRY GENERATOR's output on the structures of her own poems, several of which have been published.

Thomas A. Easton of Belfast, Me., thinks the best way to generate computer poetry necessarily involves a symbiosis between human and machine. He has written a semiautomatic program called THUNDER THOUGHT that provides ideas for poems. Again, I quote the author of the program:

"I conceive of the creative mind as having two components: the popcorn mind and the critical mind. The former generates random combinations of whatever words, ideas and images happen to be in a sort of mental focus (along with peripheral material, which is why the popcorn mind can surprise us). The critical mind then discards as garbage the vast bulk of what the popcorn mind produces and edits, twists and elaborates the remainder to form poems."

Relying on internal lists of nouns, verbs, adjectives and adverbs, THUNDER THOUGHT operates roughly like West's POETRY GENERATOR, arranging words into sentence frames to produce what Easton regards as raw poetic material for a human mind to refine.

"The intermediate result is ungrammatical, nonsensical, ridiculous garbage ... but not always. Among the many lines of garbage there always lie a few lines to which one responds. They make sense—or seem to. They beg one to tweak them a little. A pair of them insists that one make up a third line. They stimulate one to think of other lines that can accompany them.



Sonnet CXXX-b

My Apple's screen is nothing like the Sun;
The Cray is faster far on problems big:
If Apple pleasant be, th'Atari is more fun;
If wires be hairs, her circuits are a wig:
I have seen pixels dancing, red green blue,
But no such pixels see I on her tube;
And mainframes dance a logic far more true
Than in my Apple's tiny crystal cube.
I love to watch her print, yet well I know
That line printers hasten with more speed;
I grant I never saw a virtual process go,
My Apple, when it works, does in small steps proceed;
And yet, by heaven, I think my judgement sound
As any computation she has found.

An unknown author and ORPHEUS combine talents to compose a sonnet

A little editing, interpolation, elaboration and—voilà!—a poem."

Easton has written some 110 poems by this method and has published 32 of them—some have even appeared in literary journals. That ratio, he claims, is one that would turn many real poets green with envy.

The last word on computer poetry goes to ORPHEUS, a program designed by poet Michael Newman of New York City. ORPHEUS lays out strict frameworks, from haiku to sonnets, into which human writers may insert their own chosen words. Essentially ORPHEUS is a word processor ("poetry processor" in Newman's parlance) that lays out the lines of a given poetic form. The program allows a human being to fill in the lines according to whim and then to end them with the help of a rhyming dictionary. Setting ORPHEUS in sonnet mode, for example, one might write a pair of lines (in imitation of Shakespeare's 130th sonnet) as follows:

My Apple's screen is nothing like the sun;
The Cray is faster far on problems big:

Because the first line ends with the word "sun," ORPHEUS consults its rhyming dictionary and displays a number of words that rhyme with sun: bun, done, fun, gun and so forth. Scanning them, one's eye might alight on the word "fun." Is there a computer that is more fun than the Apple?

An association with games brings the Atari computer to mind for the next line.

If Apple pleasant be, th' Atari is more fun;

The first quatrain is polished off by selecting a word that rhymes with big.

If wires be hairs, her circuits are a wig:

The rest of the sonnet can be read in the illustration at the bottom of the opposite page.

The poetry programs I have mentioned may be bought by readers who want to sharpen their prosodic skills. Newman will be happy to supply ordering information for ORPHEUS, NERD (Newman's Electronic Rhyming Dictionary) and related products to those who write him at 12 West 68th Street, #2C, New York, N.Y. 10023. Easton meets West in a software package containing THUNDER THOUGHT and a program similar to it called VERSIFIER, which was written by West. Readers may inquire about the package by writing to Easton at Box 805, R.F.D. 2, Belfast, Me. 04915.

Readers may also be interested in what is perhaps MARK V. SHANEY's magnum opus: a 20-page commentary on the deconstructionist philosophy of Jean Baudrillard. That effort was directed by Rob Pike of the AT&T Bell Laboratories, with assistance from Henry S. Baird and Catherine Richards. The commentary can be obtained by writing Pike at the AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, N.J. 07974.

Students in Italy, stockbrokers in Singapore and physicians in the U.S. have joined the growing crowd of Mandelbrot-set devotees—all thanks to a ride on the Mandelbus, which I described in the February column. The effort to make the set's basic iteration algorithm understandable paid off in ridership and perhaps even readership. Yet, as a number of letters show, some confusion remains. It is only the Mandelbus' *first* stop that is being tested for membership in the Mandelbrot set. Subsequent stops may be either inside or outside the set, but if any of them turns out to be more than two units away from the origin, the first stop can be automatically excluded from the Mandelbrot set.

Readers who attempted to visit the area of the Mandelbrot set that Andrew LaMance calls Love Canal were disappointed to find empty space: they were given a wrong coordinate address. A minus sign placed in front of the first coordinate given should put readers squarely on the site of the curiously alluring pollution.

I may have been overly impressed by the magnification of 54,000 that Ken Philip of Fairbanks, Alaska, achieved in generating an image of a sea-horse scepter. Such magnification hardly marks the limit of a computer's acuity. Indeed, magnifications of that order are nearly routine both for A. G. Davis Philip of Schenectady, N.Y., and his brother Ken. The Schenectady Philip writes, "While I was in Fairbanks in November, my brother's Mac II produced a picture of a [Mandelbrot] midget at a magnification of 2×10^{31}. I consider that remarkable."

Peter Garrison of Los Angeles coincidentally explains that the double-precision mode available in most personal computers—in which the number of bits in the computer's "words" are doubled—actually more than doubles the computer's resolving power. In fact, the power is doubled for each one-bit increase in effective word size. Extremely high magnifications can therefore be achieved by means of the even greater precision made possible by special hardware and software.

Near the end of the column I mentioned a fast new algorithm for computing the Mandelbrot set that was described by Uval Fisher in the book *The Science of Fractal Images*. William S. Cleveland of the AT&T Bell Laboratories wrote to explain that the algorithm was actually developed by William P. Thurston and Allan R. Wilks. According to Cleveland, the new algorithm not only is faster but also produces more accurate pictures of the set than the standard algorithm does. As Cleveland says: "If you get on board the bus with the Thurston-Wilks algorithm painting the scene (in black and white), a wholly new and more realistic world will open up to you."

A new way to render beautiful Mandelbrot images using the traditional method was communicated to me by Carl G. Nugent of Seattle. It is now possible to see the set in delicate bas-relief, looking like the compressed fossil of an alien life-form. Although shaded in a single color, the images are just as beautiful as the full-color treatments because of their incredibly tactile nature: one can "feel" those delicate tendrils. The appearance is based on a trick that makes the tiny, intricate details of the set look from afar as though they were illuminated from one side.

The underlying idea is to divide the display screen into diagonal rows of pixels that run from the top left to the bottom right of the screen. If the iteration count generated for each pixel is taken to represent the pixel's "height," then an imaginary light source in the top left corner of the screen will cause a black "shadow" to be cast on certain pixels, depending on the height of the neighboring pixel (above and to the left) in its diagonal row. In even diagonal rows a pixel is not to be colored black unless its neighbor's height is strictly greater than that of the pixel; in odd diagonal rows, however, a pixel is colored black even if its neighbor has an equal height. Hence, in Mandelbrot "plateaus" (areas where all pixels have an equal iteration count), the diagonal pixel rows will alternate in color. Up close the displayed plateaus have a checkerboard appearance. "To make it look gray," says Nugent, "just throw away your microscope!"

FURTHER READING

COMPUTER RECREATIONS. Brian Hayes in *Scientific American*, Vol. 249, No. 5, pages 16–24; November, 1983.

BOOKS

Botany's populist, the retina in the nose, scythe and camel, life at the top

by Philip Morrison

ASA GRAY: AMERICAN BOTANIST, FRIEND OF DARWIN, by A. Hunter Dupree. The Johns Hopkins University Press (paperbound, $14.95).

Perhaps no historian of American science is better known than Hunter Dupree. This engrossing volume was launched as his dissertation in 1948, enlarged over a decade and then first published in full 30 years ago. Reissued in the centenary of Asa Gray's death, this vivid story of the man and his times so illuminates current institutions and issues around science in our country that it will enrich any general reader, even one far from garden and herbarium.

Gray grew up in a quiet green valley some 10 miles south of Utica, N.Y. The boy was an avid reader and a good student; in 1825, 15 years old and just enrolled in a local academy, he went with his new classmates to watch the splendid barge that bore Governor Clinton and his party down the new Erie Canal. By the age of 20 Asa held a New York degree of doctor of medicine and had entered into a country practice.

Along with thriving commerce and religious zeal, science too was astir in fast-changing upstate New York. Gray's own professor of chemistry and materia medica was an engaged naturalist; through his personal example and his books and collections, Gray was by 17 put under the spell of botany. The young enthusiast worked hard at seeking and ordering the local flora; at 19 he traveled to Albany (where he encountered "a grave-looking man . . . Professor Henry, who had just been making a wonderful electro-magnet") and on to New York City to call on a botanical notable with whom he had exchanged specimens and correspondence.

That was John Torrey, pioneer of American botany, who himself taught chemistry in the College of Physicians and Surgeons. "I get my bread by it . . . ," he wrote, "yet I love Botany more." He had published a well-known *Flora* and was afloat on "the currents of world botany." By the age of 25 Gray had finished with medicine and the provinces. He became a collector, helper, co-author, friend and guest of Torrey and his amiable family; unpaid for a good while, Gray enjoyed the Torreys, their lively home and busy city life.

Step by step Gray matured. He obtained support from the state and federal governments, toured and studied the collections of Europe, passed beyond field and herbarium to library and laboratory, published both monographs and texts. There was newly a market for basic books among collectors nationwide; his *Manual of the Botany of the Northern United States,* hastily written in 1848, is still in use after six generations of revisers. His output grew with the rich collections that were by-products of American continental expansion, and he entered parity with world leaders of botany after a frustrating but enlarging encounter with the specimens from American expeditions flung across the Pacific to Hawaii and on to Commodore Perry's Japan.

In 1842 he became professor of natural history at Harvard College; the institutions of American science were rapidly crystallizing. Soon thereafter Harvard appointed as professor of zoology and geology the charismatic, prolific and world-famous Swiss naturalist Louis Agassiz. He had come to America ostensibly to examine nature in the New World but stayed for life, to charm and quicken the minds of scientists, intellectuals and the powerful alike. There ensued a drama that would last a lifetime; in it Agassiz played the Hamilton of the new Republic of Science, Asa Gray a humbler Jefferson.

Should America have a powerful national academy in its capital city, to set impeccable standards outside the diverse campuses, far from the flood of students and aspirants? Should the universities themselves be graced by genuinely able Professors, one to a subject, chosen by merit for a lifelong leadership of research? Gray was something of a botanical populist. He wrote for regional gardeners and collectors; he remained in close touch and direct correspondence with a nationwide network of amateurs. Agassiz was warmly and genuinely engaged with his chosen and able students, but those who were not expert he never noticed, although he gave luminous popular lectures to wide audiences.

From his voluminous examinations of plants, Gray took a feeling for their essential variation; Agassiz, an even more versatile and tireless investigator, came to his specimens imbued with the rigid metaphysics of his own formative years in Paris and in Prussia. For him species were the special creations of the Deity, tending toward an ideal form that no individual specimen need fully exemplify.

Darwin's *Origin of Species* appeared, much heralded among the experts, in 1859. At hand as well was the fiercer turmoil of civil war. By 1863 a new National Academy of Sciences was chartered, the outcome of a famous intrigue in which Agassiz was the mainspring. Gray was more quietly at work against the high plans, however, and he and his friends won a subtle victory; the Academy remains today, modestly quasi-official, honorific and even influential, but it is no ruler over great laboratories and institutes, no Prussian or Russian Academy. By the time Agassiz died Harvard College had become a research university (it was not alone), with half a dozen professors of every science—in most cases men with complementary specialties, rarely the chosen of a school led and certified by a single titular Professor.

As with institutions, so with ideas. Gray at once concurred with Darwin: species are genetically connected, and they vary and evolve through space and time. Agassiz bitterly opposed Darwin's theorizing and use of practical evidence from domesticated forms, although he would moderate as he grew older.

Gray retired in the year Agassiz died; he wrote graciously of his late adversary. The excited schoolboy at the Erie Canal was now first botanist of the U.S. His career was one with the national expansion. The unprecedentedly wide view he had of plants on both shores of the Pacific led to his finest paper: an explanation of the

deep similarities among the flowering plants across that ocean.

"Nothing has surprised me more," Darwin wrote, than the affinities between Japanese plants and the plants of eastern North America. (Western American forms were distinct from both.) It seemed a curiosity of double creation: some species distinct but similar, some indeed identical. Gray saw the mark of a single origin, the plants long ago spread unbroken across the Bering Strait from the Carolinas past Japan and even to Europe. Split into two great branches as they drew southward from the ice, the continental flora mingled again once the ice dwindled. Change was also clear. "Sometimes, and at some places, one species did turn into another," Dupree remarks.

"Behind Gray stood the tradition of empiricism and of medieval nominalism," the old understanding that the names affixed by mind are not intrinsic. "Behind Agassiz stood romanticism, transcendentalism, ... Goethe," all sharing the rival medieval stance: reality seen as essentially conceptual, a mirror for the realm of the ideal. In 1859 "the prize included ... man himself"; so it does still.

Gray was a deeply religious man lifelong, a serious and consistent "orthodox Presbyterian as my fathers were." He saw no conflict between his evangelical faith and the evidentiary rationalism of the Enlightenment. The Holy Bible was not divine revelation, but the human record of that revelation, and it was no textbook of science. Where the evidence was clear, he felt, "settled scientific belief must needs control" religious beliefs "or what were taken for such." He proclaimed his own strong faith in the old Nicene Creed that states simply "I look for the resurrection of the dead" and does not, he believed, entail "the crude notion of the revivification of the human body." Such a doctrine required no dispute with Darwin, although Gray hoped rather wistfully that evidence of design might yet appear, perhaps in the unknown mechanisms of descent and variation.

On Agassiz' grave in Mount Auburn Cemetery stands a great glacial boulder brought from an Alpine meadow; on Gray's there is a simple headstone with a cross, marked "Asa Gray 1810–1888." This reviewer was fascinated by the democrat, but the grandee too was prepossessing. For him seek out *Louis Agassiz: A Life in Science,* by Edward Lurie. It is a brilliant social biography: less science, more fireworks. Lurie's volume was also born in the 1950's as a dissertation and lately reissued, a fit companion to Dupree's *Gray,* in the same bargain format.

PERFUMERY: THE PSYCHOLOGY AND BIOLOGY OF FRAGRANCE, edited by Steve Van Toller and George H. Dodd. Chapman and Hall ($49.95).

The chemical sense is the deepest and the least understood of all our sensory channels. A small patch of some 25 million ciliated, mucus-covered and constantly renewed smell receptors lies at the top of each nostril passage. They report to the olfactory bulbs on the undersurface of the brain through unusually convergent wiring, many receptors mapped to a single secondary cell. Fewer than a dozen molecules suffice to excite a receptor; 40 excited cells can induce the sensation. All of this recalls the retina of the eye (read "molecules" for "photons" throughout). Only in the 1980's has a critical mass of biochemists joined in a study of the channel. By now there is a basis of empirical evidence on receptor events.

The retinal analogue is explicit, if unproved: there are powerful inhibitory as well as excitatory odorants; multiple binding sites of odor molecules to a variety of protein receptors can be the combinatorial source of remarkable specificity; signal amplification arises from enzyme shifts that modify reaction rates and thus the concentrations of a few specific messenger molecules. Add the clinical existence of odor blindness and even of hallucinations (often the consequence of head injuries), and the sense of smell begins very slowly to take its place in molecular neurobiology.

All that is said in this book, but to report so blandly would badly misrepresent the burden of this unusual symposium. The Olfaction Research Group of the University of Warwick, where the editors work, has assembled papers by biochemists, perfumers, marketing researchers and even psychotherapists. The First International Conference on the Psychology of Perfumery was held there in mid-1986. The papers center on our species, with the five parts grouped around the biology of scents (from evolutionary to molecular); the social aspects; new electroneurological findings; fragrance therapy; and the consumer and the seller of perfume.

The powerful evocation of memory by scent has been as famous as it is familiar since Proust entered on his remembrance of things past through the orange fragrance of a teacake. The direct neural links between the olfactory sense and the limbic system involved in emotive behavior (which was once called the rhinencephalon, or nose-brain) have long been read as a sign of the antiquity of the sense, antedating our primate cortex. The external hormonelike substances called pheromones that are decisive for the behavior of many organisms are suggestive of that antiquity.

Certainly there is a tension here, close to paradox: if the sense is so old, so strong in affect, why is the human culture of smell so impoverished compared with those of sight and hearing? Sacred incense and personal perfumes and odors are about all that culture holds of scents. (The art of the dining table might be given some weight here, but its richness does not, in fact, focus on odor.)

"The most ... expensive modern perfumes contain a number of mammalian sex attractants," and trace secretions in human sweat contain odorous steroids found in the testes of pigs. Yet in infants odor associations seem to be learned; the transient event is all-important, the nature of the odor arbitrary. Trygg Engen of Brown University even writes that "the olfactory system is a *tabula rasa.*" Somehow, by nature or nurture, or by both, human innate response to olfaction has been largely repressed. We are visual primates who have lost our card at the world's library of scents. There are a few old books of scent we still like to read, but their once thunderous mandate has moderated into a still, small voice. Perhaps we remain especially sensitive to certain ancient molecules. "It is difficult ... to think of a successful perfume which does not contain" some molecule comparable to musk and hence related to human trace odorants. The issue is debated here in several very readable papers.

The unorthodox therapists who contribute here are herbalists, implausibly vitalists: for them natural essential oils are somehow distinct from synthetic ones. The marketplace does not concur. Ever since Chanel No. 5 the most successful novel perfumes have included major synthetic odorants in their complex recipes. The nose is also the seat of a second sensing system, the trigeminal, mediated by a quite distinct nerve pathway to the brain. That system responds especially to such pungent vapors as ammonia; the head-averting reflex elicited by smelling salts is not strictly due to smell.

Current fashionable perfumes include some remarkable synthetic substances. For example, the compound beta-damascenone is found at one

part in a thousand in natural rose oil. The fragrant molecule enjoys an odor threshold at a dilution in water solution down to only seven parts per trillion, unmatched by any other substance in the rose, and that fraction may well dominate the combined fragrance of all the rose's 275 known constituents. Today synthetic beta-damascenone is the "key impact" ingredient of new perfumes that at least marginally act by stimulation of the trigeminal system. The toxic brand names and startling advertising are almost as far outside earlier experience with scents as are their ingeniously synthesized molecules.

INVENTION AND EVOLUTION: DESIGN IN NATURE AND ENGINEERING, by M. J. French. Cambridge University Press ($59.50; paperbound, $22.50).

A scythe is a superb and ancient tool marred by what seem offhand to be visual oddities: the angles of blade and handles appear to miss the appropriately symmetrical planes. Yet the entire form arose from two millennia of beautiful evolutionary adaptation. The awkward-looking offsets function to help transfer the reaper's weight as he swings from one foot to the other and to allow the cut crop to fall in a neat swath.

Nor is the comical-seeming camel the result of a horse design botched by a committee, as the old quip goes. No horse can survive in places where a camel thrives and may even travel at a rhythmic gallop. The desert animal is "an elegant design of the tightest unity." Those padlike feet, the knobby knees, the hump for storing food and the lippy profile "have a congruity that derives from function . . . a certain bizarre elegance," not found in the feeble "undesigns" of the monsters of heraldry and legend, untested by the needs of reality. The *Winged Victory of Samothrace* is an exception, beautiful even though the Greeks "scarfed together incongruous parts" to make a simulation of a functional system, however inadequate for flight her tiny wings may be.

The author, professor of engineering design at the University of Lancaster, is a man of independent taste. But he is no naysayer; he is an honest enthusiast with an aesthetic he is able to explain. The pick-up cartridge of a record player "may marry a tiny precisely-shaped diamond, . . . wire fine as spider's web, . . . a magnet . . . of rare metals or oxides, . . . unsuspected a century ago. . . . all this is offered, not to kings or bankers, but to any citizen. . . . For . . . half a day's pay he can buy this triumph of craftsmanship and ingenuity, beyond anything made by Fabergé." (That was written a few years back; even stronger is the case for the delicate jewelry of the infrared laser scanners in a CD player.)

The text is evenhanded as between the stepwise engineering done by human beings and the patient opportunism of organic evolution. Analogies abound. Fish and naval ships, unlike workaday freighters, need to provide for a severalfold peaking of power for high-speed dashes. The warship commonly has both cruise and main engines; the fish has both red and white muscles. The breast meat of chicken tells of the bird's origin in a ground-dwelling species: the white muscle is glycogen-fueled, suited to the short power bursts of brief escape flights, whereas the dark myoglobin of chicken legs can oxidize fats on the spot for sustained power output.

More than half of this text, with plenty of telling numbers, tables and graphs, but few equations, is a designer's-eye look for what the designer needs from basic physics. Chapter after chapter reviews forces and energy, materials, mechanisms and structures. The air of purposeful attention and the wealth of both familiar and curious examples give this overview a freshness that will do many a reader good, even a sophisticate.

Soon enough the view turns closer to design itself, its practice and philosophy. True to the author's independence of mind, the systems-with-everything approach wins faint praise. Its generality renders it sloganlike, although its way of avoiding the mental "grooves of past practice" is valuable. Our hunting forebears might have worked on a stiffer bow, but what they really were after was "a better deer-catching system." Even discounting traps and the like, "camouflage and deodorants to help in stalking" might have been more strategic. In any case, in the end it was domestication that brought us more meat—perhaps to threaten health?

There is a fine account of the cumulative nature of design, illustrated by the vast repertoire of available bearings and pivots. One single good-sized firm in the U.K. offers 10,000 different bearings (of all sizes and materials) but by no means the full range that is out there. Pumps and pressure vessels are similarly discussed, with a nice example of how the device can be characterized by a few key quantities that place each possible design at some point within a graph field. De-

vices from windmills to steam turbines can be mustered there according to relative blade speed and flow volume, cleverly made to be comparable over a wide range of sizes. Here one senses the style of well-founded engineering practice.

A similar fascinating chapter treats production and design. In nature, of course, we call it reproduction: blueprints, jigs, tools and dies become DNA-coded templates. Evolutionary similarities are easily found. Still, engineers are more calculating today, freer from what went before. Deliberate and confident innovative design is more frequent, despite market and testing lags. "Nevertheless, . . . the evolutionary aspects of design are declining . . . . Evolution itself is becoming extinct."

The book ends with an analysis of design training. The discussion of creativity in the language of convergent and divergent thinking is not very helpful for design; a creative designer often converges from a wide base of the potential onto a narrow but admirable choice. The strongest position argued here is the contrast between visual and verbal thinking. It is all but impossible to design in detail with words alone. You can easily speak of "dog" in general, but to draw one is hard. The concreteness of drawing is evident, and design must pass beyond abstraction to the concrete. French has no doubt that both modes of thought are important. "Homo jabber . . . ," he writes, "thinks he is no end of a fellow, and looks down on humble homo faber, the engineer and the designer." It is wordsmiths who manage governments and schools.

A faint echo of Thorstein Veblen may be heard in this last chapter, concerned as it is with setting engineer and designer within society. But this readable, original text ends very quietly; the wisdom we need in this epoch is not yet in the engineer's hand.

HIMALAYAS: GROWING MOUNTAINS, LIVING MYTHS, MIGRATING PEOPLES. Text by Blanche Christine Olschak, Augusto Gansser and Andreas Gruschke, concept and design by Emil M. Bührer. Facts on File ($40). **HONEY HUNTERS OF NEPAL,** by Eric Valli and Diane Summers. Harry N. Abrams, Inc. (paperbound, $29.95).

These two large and splendid translations could be called coffee-table books, save that the beverage would be out of place: it is a tea tray they should grace. They have the sheaf of showy images of travel and the personal text of the genre, but their con-

tent is so rewarding and original that they invite study even as they make it attractive.

Early in the book about the Himalayas, a Swiss and German collaboration, one photograph shows a fine ammonite fossil, big as two hands. The Tibetans hack them from the black Jurassic/Cretaceous slate of the valleys and carry them as amulets—the spirals evoking the eternal wheel of life—up to the high passes, where they deposit their specimens in "quite the wrong surroundings," to the confusion of the geologists. Later in the volume you encounter the gilded face of a big modern image of the Buddha in Ladakh; on the serene wide brow his third "eye of wisdom" appears plainly—in the form of an ammonite.

The entire book is a fresh visual essay on the "spirit of place," the close coupling between the complex human cultures that dwell within the greatest mountain range on earth and the heroic theater of rock, ice and water in which they thrive. Half of the text is the work of veteran geologist Gansser (who carries the high title of Baba Himalaya, Father of the Himalayas, granted him by the University of Peshawar), whose insight and lengthy travels take us into past and present, heights and depths. Most of the rest—architecture, pilgrimage, myth, everywhere an exacting piety in celebration of the divine—is set out by Tibetologist Olschak. Finally an appendix offers a reference gazetteer and history of the states of the region.

Orbital photographs and fine maps, both topographical and geological (one colorful map in Chinese), support the story of the drifting continental plates that still both heap up and dig under the great mountains, all the way north to the broad Tibetan valleys.

One double spread displays the pride of the range, the holiest of peaks, Mount Kailas, breathtakingly white and as symmetrical as a "crystal stupa." It is the literal center of a pilgrimage, for its lonely peak is circumambulated by many Tibetans and Hindu pilgrims from India, trekking for three days through the stony passes that surround it at 18,000 feet. The most devout mark their way by prostrating themselves headlong, the full length of their bodies, over and over again along the road. In contrast, the old and the ill may hire proxies and yet claim the meritorious circuit; it used to cost one sheep. Sven Hedin made the trip arrogantly on horseback in 1907, and Gansser himself completed his geological round in 1936, disguised as quite another sort of pilgrim, for "the very stones of this region are sacred": no hammers. Fields of blue-violet aconites, bluer holy lakes and medieval ruins are seen in splendid photographs; the traveler tells of the piles of human hair, the forest of cairns, the newly extracted teeth that mark the pilgrims' votive offerings, obedient to the teaching that one circuit would erase the sins of a lifetime.

Mount Kailas is unique both in form and in site. Within a 50-kilometer circle around its flanks rise the three greatest rivers of these lands: the Indus, the Brahmaputra (called Yarlung Tsangpo in Tibet) and the Ganges, as well as two major tributaries. The peak itself is truly Asian, not the gift of drift from the polar south.

The text chapters take us down each Kailas-born river to tell its wonders. The extraordinary geography leads to equally dazzling works of the hand. Most readers will be surprised by what is shown and by the knowing ethnographic and historical context given. Stunning examples are not few. The dust jacket repeats one surprising scene: rounded granite boulders the size of houses—true gray elephant rocks—lie on the track to a monastery not far from Lhasa. The site is impressive and yet hardly unique, but here the monks have long painted and renewed large, loud polychrome paintings of their gods and demons. Again: in a small, sacred pond near Kathmandu rests the god Vishnu. He reclines at his ease, a wheel in one spare hand, atop the dozen intricate coils of the serpent Ananta, symbol of "never-ending time." Cool waters lap around the two beings carved smoothly from a single block of what looks like basalt. The King of Nepal is forbidden to visit this image of his forebear.

In the thin air, graffiti are sacred. Looking for all the world like some work of Robert Indiana transliterated, a set of big carved letters rendered in strong primary colors on a white-painted ground embellishes a rock wall, one of innumerable examples of the "sacred syllables of the creatures of the six worlds," squeezed by past European scholarship to yield (incorrectly) the Sanskrit phrase about a jeweled lotus. An austere inscription decorates in dignity the entire face of a huge boulder out in Sherpa country, a white granite intruded after the mountain-folding stopped.

The second book takes us to the foothills of western Nepal, to where the Gurungs live. Once this folk survived as hunter-gatherers; now they mostly farm their sparse land. The prosperous among them still serve for pay and pensions in the famous Gurkha infantry regiments; some older men remember El Alamein and the wines near Monte Cassino. But one hunter plies his specialized form of the chase. Hanging on a long rope ladder halfway down a sheer 450-foot cliff, he subdues by smoke a hive of wild bees, to reap and sell the wax of their crescent-shaped comb and its 50 pounds of golden sweetness. The daring photographer rappelled down the cliff beside Mani Lal, to photograph him close-up within the cloud of smoke and angry, stinging bees. In dozens of color photographs, many reproduced at unusually outsize scale from the sharp Leica shots, we meet and come to know and admire all the hunting party.

"No one can remember how long we have been eating the honey of the cliffs," says the honeyhunter. The authors, French ethnographers, have documented with admiring intimacy today's search for honey in the rock, and they also show us accurate representations of the technique in cave paintings 12,000 years old, to be found on three continents. But Mani Lal, now 63 years old, has decided to give up the harvest of the autumn honey. "With him, this timeless quest will end, for he has no successor."

Vishnu recumbent on the serpent Ananta: a photograph from Himalayas

THE LIFESTYLE RESOURCE™

TO YOUR HEALTH

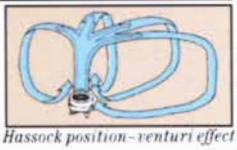
Ancient mandarins dating back 800 years believed that these Chinese Exercise Balls induced well-being of the body and serenity of spirit. These treasured gifts were given to President Reagan and his wife while visiting China. The Chinese say that rotating the balls in the palm of each hand stimulates the fingers and acupuncture points, and improves circulation of vital energy throughout the body. Sports enthusiasts, musicians, computer users and health-conscious people everywhere consider them great muscle conditioners. Arthritis sufferers feel a decided benefit from this gentle but challenging exercise. Very effective for relaxation and meditation, Chinese Exercise Balls emit a distantly mysterious chime as you turn them. Beautifully handcrafted, 45 mm. hollow polished chrome balls are perfectly weighted and fit comfortably into the average man's or woman's hand. In exquisite silk brocade box. **$29.95** #1701.

A WATCH FOR ALL REASONS

This handsome digital-analog chronometer will swim with you (to a depth of 150 feet), fly with you, calculate your speed, wake you in the morning, and tell you the day and date. A push of a button will convert the digital display to 24-hour timekeeping. All functions—digital and analog timekeeping, calendar, alarm, chronometer and tachymeter—are powered by a highly accurate electronic quartz movement. The analog hands and hour markers are luminous. The stopwatch times to 1/100th of a second, and the rotating bezel can be used for such practical matters as telling you when the parking meter will need another quarter. Adjustable fit. **$39.95**. Teflon coated durable black matte anodized stainless steel #1051; Stainless steel and gold-plated #1061.

AERODYNAMIC COOLING

Turbo-Aire's jetstream



Hassock position - venturi effect

The Turbo-Aire fan moves more air and provides more cooling than ordinary fans - you get more than twice the air delivered by oscillating room fans. Patented and computer designed to maximize efficiency and minimize noise and vibration. This fan creates an exceptionally strong, smooth jetstream column that sets a room stirring with secondary air currents. Refreshes better than the intermittent blast of air from an oscillating fan. Aerodynamic housing increases blade-tip efficiency over conventional fans. Floor, table or wall mount, adjusts to any angle. Set in the hassock position, the five bladed, 12" fan redistributes air in an entire room, making it as useful in winter as in summer. Can reduce air conditioning costs and is 300% more energy efficient than ordinary fans with the same type and size motor. With a 3 speed fan control, the Turbo-Aire is the most advanced 12" portable house fan. Weighs less than a case of cold soft drinks - and more easily moved! Manufacturer's 2-year limited warranty. UL listed. **$79.95** #2161.

AN AMERICAN CLASSIC

The *very best* of the Adirondack Chairs are still made in the Adirondacks, in Willsboro, N.Y.. Willsboro's skilled woodworkers have done more than just re-create the original Adirondack Chair - they've added refinements such as a fully contoured seat and 6" wide arms for the ultimate in comfort. And the workmanship! The seat and backrest are hand-dowled and glued to the frame for strength, using solid hardwood pegs (not capped screws) in the true Yankee manner. The rich rock maple is an extremely dense and attractively grained wood, all sanded - no sharp corners or splintering. A unique (and patented) tongue-in-groove folding mechanism lets you fold the chair to a mere 10" thickness for storage. Each completed piece is treated with a wood protectant, ready for you to apply a clear finish or paint, or leave it to weather naturally. There has never been an Adirondack Chair like this...and we have a matching footrest, designed to create a chaise lounge and an end table to go along side - both constructed in the same fine tradition. The chair is 38" x 14" x 20", **$139.95** #1881. The footrest is 14" x 20" x 22", **$79.95** #1891. The end table is 16" x 14" x 24", **$69.95** #1901. Some assembly required.

THE LIFESTYLE RESOURCE™

FOR FASTEST SERVICE — CREDIT CARD ORDERS CALL TOLL-FREE 24 HRS A DAY **800-872-5200**

DEPT. SFAF09; 921 EASTWIND DR. SUITE 114: WESTERVILLE, OH 43081

SEND TO (PLEASE PRINT)

ADDRESS

CITY STATE ZIP

☐ CHECK OR MONEY ORDER ☐ MASTERCARD ☐ VISA ☐ AMEX SIGNATURE

ACCT.# EXP. DATE

ITEM NO.	QTY.	DESCRIPTION	ITEM PRICE	TOTAL PRICE
			SUB TOTAL	
			SHIPPING (See table at left)	
			TOTAL	

Shipping Charge covers UPS, handling and insurance for guaranteed delivery.
UPS Second Day available for an additional $7.50 per order.

Up to $20	$ 3.95	$50.01 to $60	$ 7.95
$20.01 to $30	$ 4.95	$60.01 to $70	$ 8.95
$30.01 to $40	$ 5.95	$70.01 to $100	$10.95
$40.01 to $50	$ 6.95	Over $100	$12.95

Canadian residents please call (614) 794-2662 for ordering information.

ORDER WITH CONFIDENCE

- *Most orders ship within 48 hours of receipt.*
- *Credit Card orders billed only upon shipment.*
- *No risk 30-day return privilege.*

UPS 2ND DAY AIR

THE MEDIA DEVELOPMENT GROUP

©1989 C46 ADVERTISEMENT

ESSAY

Stages of evolution and their messengers

by Edward Rubenstein

How does nature encompass and mold a billion galaxies, a billion billion stars—and also the earth, teeming with exuberant life? New insights into how nature operates come from parallel advances in particle physics and in molecular biology, advances that make it possible to examine fundamental physical and biological processes side by side. The resulting stereoscopic view deep into the past reveals a previously hidden, unifying logic in nature: its paradigm for construction.

What nature does, in essence, is to make assemblies. It relies on the same template of programmed action—involving intermediaries, or messengers, to transmit the needed information—in each step of assembly along the way. Continuing sequences of assemblies are the veins of evolution. Biological evolution is the result of natural selection operating on random variations. Physical evolution is a similar process of construction: a chain of chance associations from which new structures arise. Whether these objects survive or vanish depends on their environment.

For instance, elementary subatomic particles, driven into motion by thermal agitation, may collide. If the attraction between them exceeds the momentum of their recoil, they may adhere. Their course as a union depends on the fitness of the combination, that is, on whether the strength of their mutual affinity exceeds the disruptive effects of their surroundings. If the attractive interaction is strong and the ambient conditions are mild, the composite is suited to its environment: if it is stable, it may have a long or a virtually infinite lifetime. Such an aggregate may engage in higher-order assemblies, whose fate depends in turn on their fitness. In this way prebiotic evolution led to the elaboration of structures characterized by increasing diversity of conformation and versatility of action, until eventually assemblies of great elegance arose that were suited to serve as the modular units of life.

Recurring themes unify the events that transpired in the early universe with events that shaped living organisms billions of years later. Physical evolution and biological evolution are both characterized by common descent, random interactions, natural selection and the eventual—and apparently inevitable—expression of symmetry. These reiterations suggest that physical evolution and biological evolution are closely linked and that they create their assemblies through common mechanisms.

In the physical realm, operations arising from the interplay of four forces are transmitted by messenger particles to induce the assembly of structural particles—the fermions—into atoms and molecules. In the biological realm, operations arising from the information encoded by the sequence of four bases in DNA are transmitted by messenger molecules to direct the assembly of structural molecules—the amino acids—into proteins that make up living cells. This correspondence reveals a fundamental program of nature: information, embedded in a source, directs assemblies through intermediaries, or messengers.

What are the messengers and the messages of each stage of evolution? The messengers of physical evolution are the intermediary subatomic particles belonging to the class of bosons. In the sense that any number can occupy a point in space-time, bosons are dimensionless. They transmit information in the language of the forces of nature, and their messages are simple imperatives: Move.

The messengers of biological evolution are the molecules of messenger RNA. They carry information in the one-dimensional linear language of the genetic code, and their messages direct the assembly of the molecular construction units of life. The messengers involved in the evolution of multicellular organisms include a large number of chemical mediators and electron-transfer molecules. They communicate information from cell to cell in the three-dimensional language of stereochemistry and neural networks, orchestrating a broad range of biological activities, sharpening perception and shaping response.

The messengers of cultural evolution are the symbols of acoustic, graphic and body language. Their messages enable individuals and groups to communicate emotions and nondimensional abstractions such as ideas and to create a new kind of assembly: a body of knowledge. By these means human beings endow their successors with their collective experience as well as with their genes.

Evolution invents. Dimensions and time arose out of chaos, physical evolution led to biological evolution, biological evolution led to cultural evolution and cultural evolution has now led to directed evolution. Henceforth, life no longer evolves solely through chance mutation. Humankind has begun to modify evolution, to bring about nonrandom, deliberate changes in DNA that alter living assemblies and create assemblies that did not exist before.

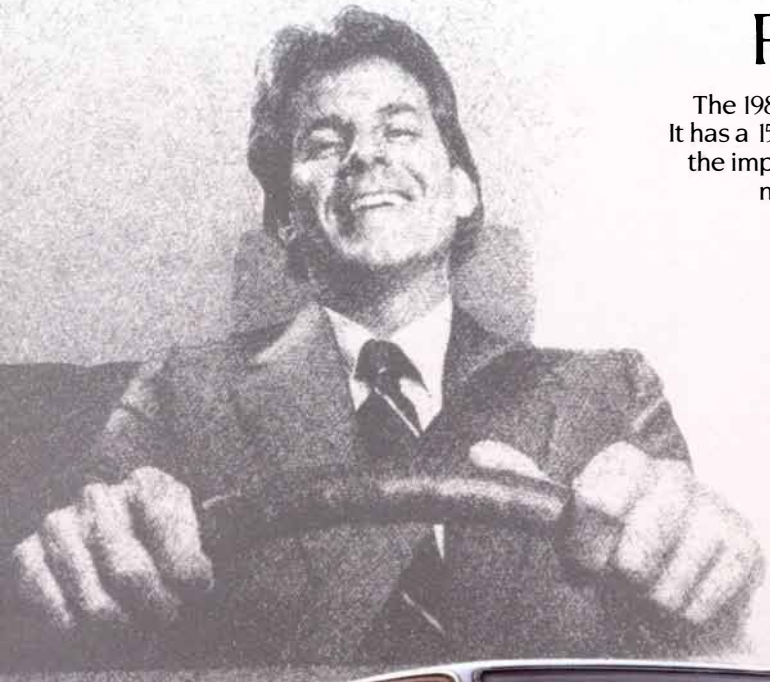
The messengers of directed evolution are human beings. Their messages, expressed in the language and methods of molecular biology, genetics and medicine and in moral precepts, express their awareness of human imperfections and reflect the values and aspirations of their species.

Human beings, then, have become active participants in the process of evolution. They make decisions about what constitutes a genetic disorder, and they have set out to detect and repair faulty genes. New problems—bearing on ethical, legal, medical, environmental, economic and social issues—will arise. Human beings make mistakes, but so do random mutations. Humans will make choices on the basis of the sway of the evolutionary past. Their role in directed evolution is defined by the fact that nature has always depended on messengers to carry out its assemblies.

The epochs of evolution telescope ever more quickly. The duration of the protouniverse (that which preceded the familiar universe) is presumably infinite; the duration of physical evolution is about 15 billion years, that of biological evolution about three billion years and that of cultural evolution about 50,000 years. Although directed evolution has just begun, extrapolation suggests its span will be brief, and so it is not too soon to wonder what will succeed it. Perhaps the increasing tendency of humans to cherish all the creations of nature, living and nonliving, and to strive for harmony with them foreshadows the age that will follow directed evolution. Who will be the messengers, and what will be their messages?

EDWARD RUBENSTEIN is associate dean for postgraduate medical education at the Stanford University School of Medicine.

ALL YOUR INVESTMENTS SHOULD PERFORM THIS WELL.



The 1989 Toyota Camry V6 will give you dividends at every turn. It has a 153-horsepower four-cam 24-valve V6 engine that delivers the impressive passing power and extra leverage you require to negotiate on the street most effectively. With all the handsome styling, roomy comfort, smooth handling and reliability you expect from the Camry name. The 1989 Toyota Camry V6. A long-term asset loaded with instant gratification.

A **36-month/36,000-mile** basic new vehicle limited warranty with no deductible and no transfer fee applies to all components other than normal wear and maintenance items.

Call 1-800-GO-TOYOTA for more information and the location of your nearest dealer. Get More From Life...Buckle Up!

TOYOTA QUALITY
WHO COULD ASK FOR ANYTHING MORE!

SEAT OF GRANDEUR.

For gift delivery of Grand Marnier® Liqueur (except where prohibited by law) call 1-800-243-3787.
Product of France. Made with fine cognac brandy 40% alc/vol (80 proof). ©1989 Carillon Importers, Ltd., Teaneck, NJ.