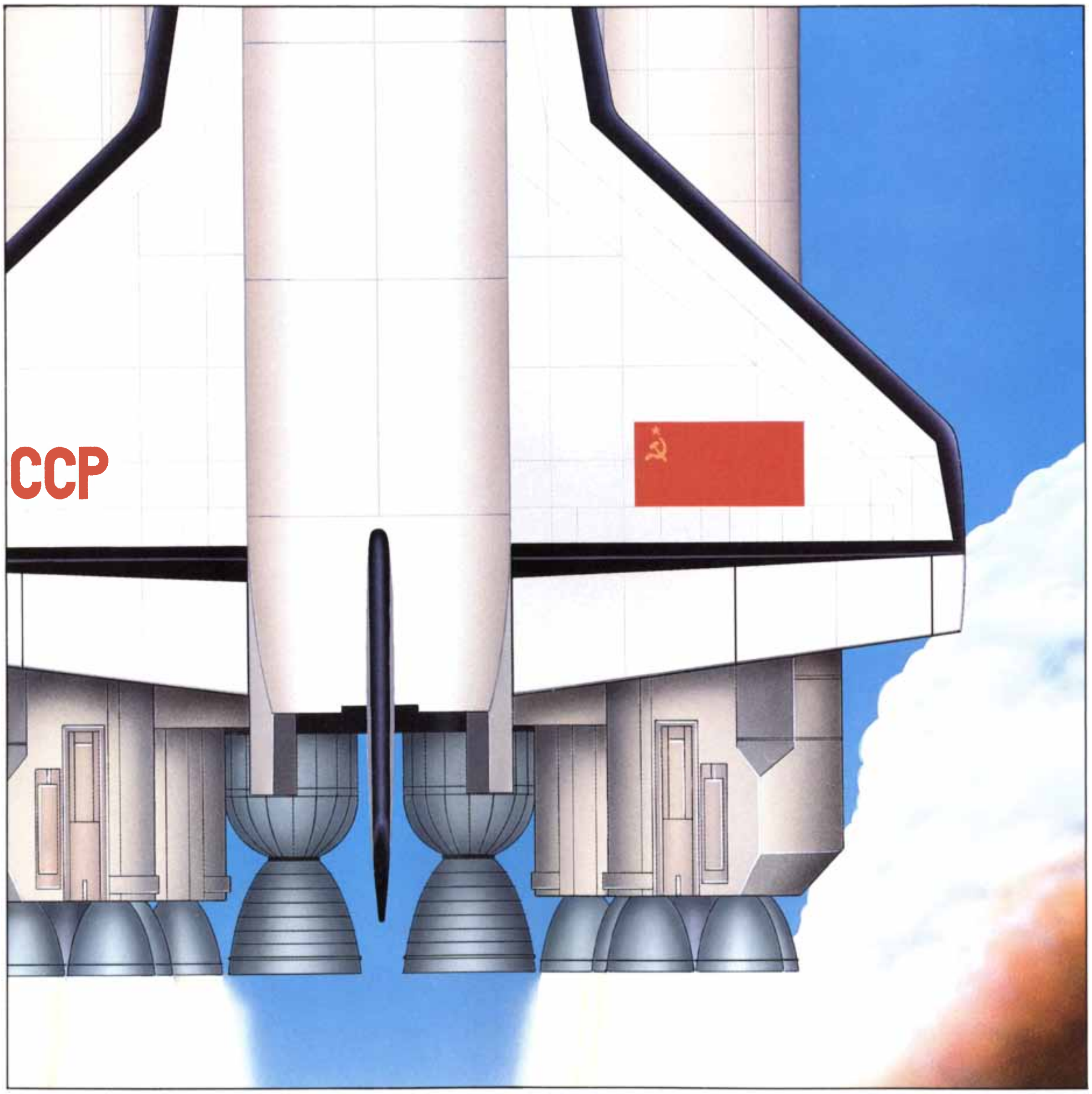
SCIENTIFIC AMERICAN

FEBRUARY 1989
$2.95

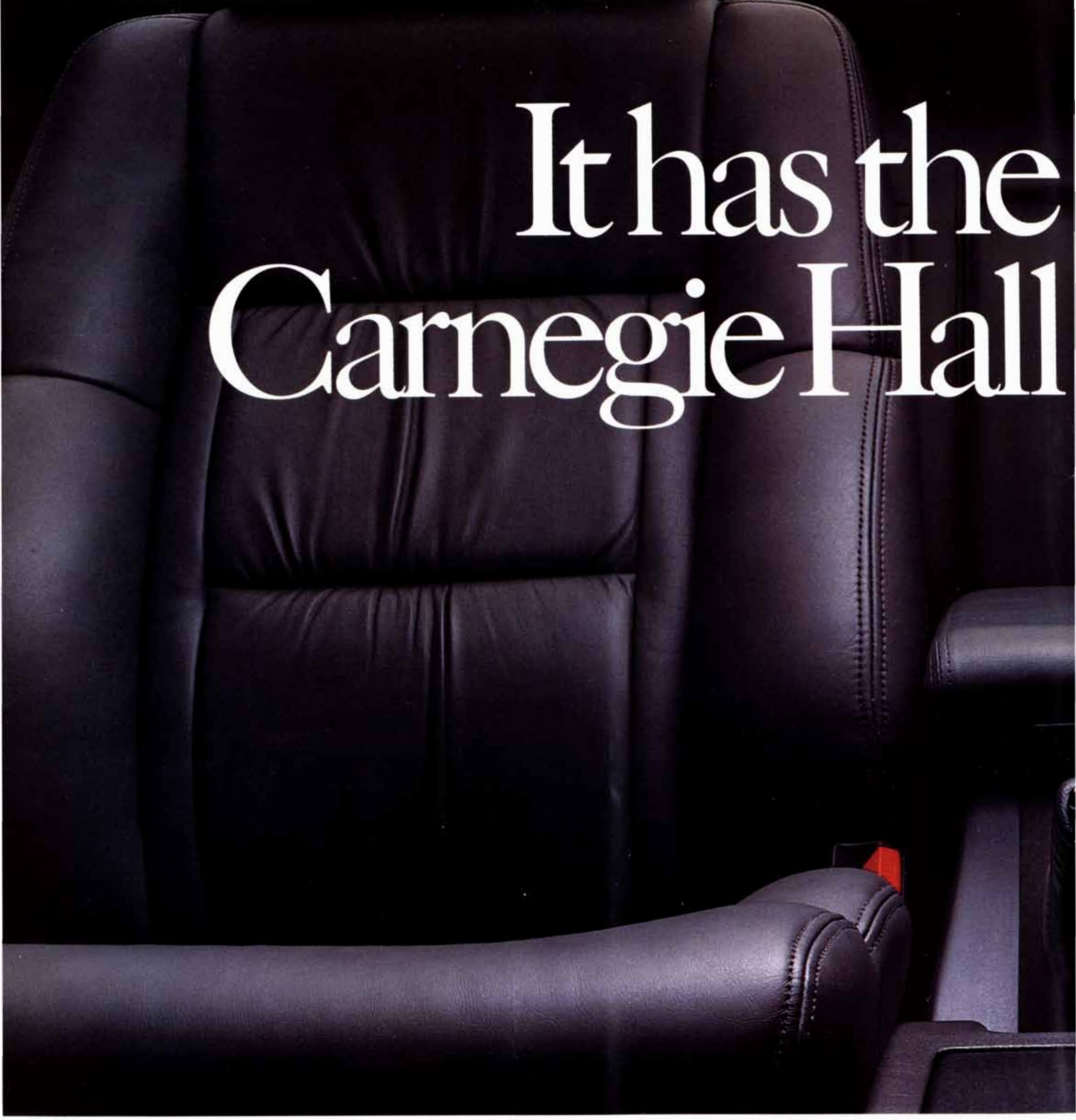
An American astronaut reports on Soviets in space.

Underwater cascades that dwarf Angel Falls.

A cool look at high-temperature superconductors.



***Soviet space shuttle:** the most recent addition to the U.S.S.R.'s thriving space program.*



This brief course in music appreciation is brought to you by the Accord SEi 4-Door and its ear-tingling Bose stereo sound system.

Important to any music system is the acoustic environment around it. The Accord is quiet. Hushed. Even at highway speed, interior decibel level is lower than the sound of a page turning.

An ideal music system is designed for the specific room in which it will be played. In this case, the Accord interior.

It started with some careful listening by a couple of audio robots. With sensitive microphones in their ears, the robots were able to tell engineers exactly what a human ear would hear from any seat in the house.

The engineers then mapped out a perfect pattern of sound. Direct and reflected. They placed speakers, amplifiers and equalizers in optimum position.

Then tuned everything.

The result is an incredibly natural sound, as if hearing a "live" concert hall performance.

So listen to an electric guitar wail through a finger-flying riff. Blast the cannons of the 1812 Overture. Or sing the blues way down low. With clarity.

Ready, maestro? A remote control on the steering wheel puts volume and station selection at your fingertips. There's also a compact disc connector on the dash.

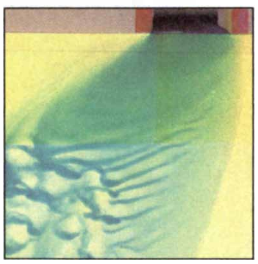
The Accord Special Edition is in tune with all your senses. The driver's seat is contoured and reclines. The cushioned leather is seductive.

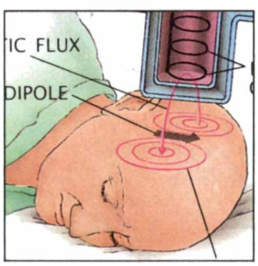
And just like front row center seats at Carnegie Hall, availability of the Accord SEi is limited.

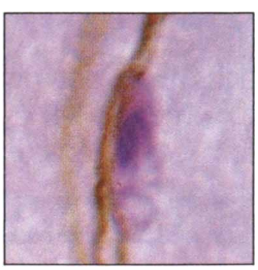
HONDA

The Accord SEi

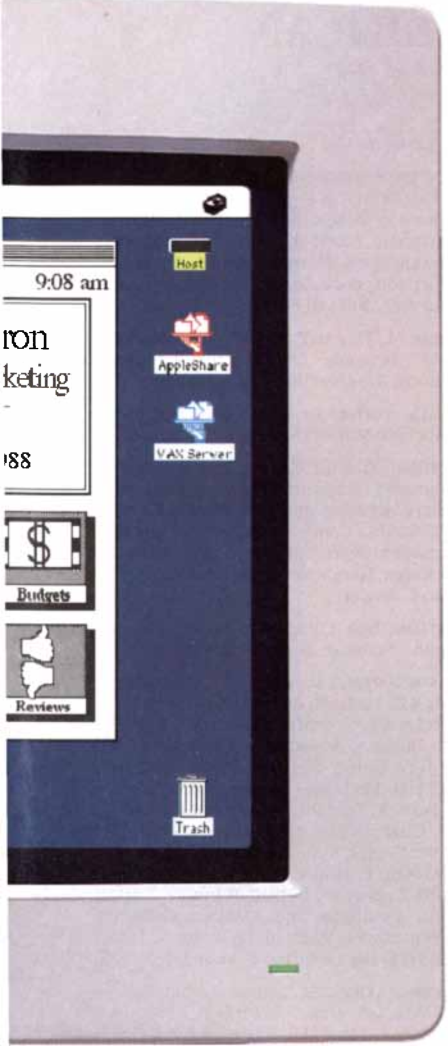
SCIENTIFIC AMERICAN
February 1989 Volume 260 Number 2

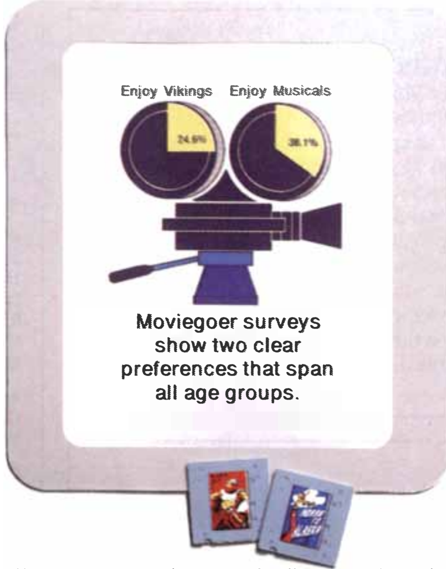






DEPARTMENTS





All your creations (charts, graphs, illustrations) can be copied into presentation pieces with typical Macintosh ease. The result being you can generate overheads, slides, leave-behinds — everything but yawns.

Reach clear to the other side of the world. With an AppleFax™ Modem, you can now send a document directly from a Macintosh to any office equipped with a fax machine — a graphic improvement in the way you communicate.

Macintosh is designed to work the same way you do. So you can more easily produce the one thing every business demands: results.

could do that.”

Or that printing a document to a fax machine in Toyohashi is as effortless as printing to the LaserWriter® one desk away.

And since Macintosh is Macintosh, such pleasant discoveries don't require months of computer training. Only common sense and an appreciation for uncommon results.

To make a few discoveries of your own, visit an authorized Apple dealer. (Call 800-446-3000, ext. 400, and we'll be happy to give you the name of one nearby.)

Before long, you'll be talking like a seasoned Macintosh pro. And saying things like:

“Anyone can do that.”

The power to be your best.™

THE COVER painting depicts the Soviet space shuttle *Buran* as it is being launched by the U.S.S.R.'s powerful new rocket, Energia (see "Soviets in Space," by Peter M. Banks and Sally K. Ride, page 32). The Soviet shuttle, which made its debut last November, is similar in design to the U.S. shuttle, but unlike the U.S. spacecraft it does not have to be piloted. Such flexibility has typified the Soviet approach to human activity in space through two decades of piloted missions.

THE ILLUSTRATIONS

Cover painting by Hank Iken

Page	Source
33	Tass/Sovfoto
34-38	Ian Worpole
39	Australian Department of Defense
43	Bruce Coleman Inc./ Nicholas deVore III
44-49	George V. Kelvin
51	Robert Frazel, Woods Hole Oceanographic Institution
52-56	George Retseck
57	Robert Frazel, Woods Hole Oceanographic Institution
60	S. Varnedoe
63-64	Thomas C. Moore
65	Johnny Johnson
66	Thomas C. Moore
67	Thomas C. Moore (*top*), Argonne National Laboratory (*bottom*)
68	AT&T Bell Laboratories
75	Fernando Nottebohm
76-79	Patricia J. Wynne
81	Lawrence A. Crum, University of Mississippi
82-85	Andrew Christie
86	Dominick J. Casadonte and Frank J. Scheltens, University of Illinois at Urbana-Champaign (*top*); Stephen J. Doktycz, University of Illinois at Urbana-Champaign (*bottom*)
89	Christy G. Turner II
90	Christy G. Turner II (*top*), Tom Prentiss (*bottom*)
91	Christy G. Turner II (*top*), Tom Prentiss (*bottom*)
94-95	Tom Prentiss
96	Christy G. Turner II (*top*), Tom Prentiss (*bottom*)
99	Quesada/Burke
100	Hank Iken
101	Quesada/Burke
103	Quesada/Burke
104-107	Michael Goodman
108	Edward Bell
109	Walter S. Strickler (*top left*), John Dewey Jones (*top right*), Andrew LaMance (*bottom left*), Ken Philip (*bottom right*)
110	Michael Adler (*left*), R. Terry Sproat and Rick McCauley (*right*)
111	James L. Crum

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Copyright © 1989 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Subscription rates: one year $24, two years $45, three years $60 (outside U.S. and possessions add $11 per year for postage). Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51593.

SCIENTIFIC AMERICAN

Established 1845

Pure brilliance.
A fantasy in crystal and silver.

© 1989 FM

In the tradition of great crystal sculpture...an original work of art by the internationally-renowned sculptor James Carpenter, whose works have been exhibited at the Smithsonian and leading museums throughout the world. Each piece is personally hand-signed by the artist. Like a mountain of sparkling ice, your sculpture is alive with the magic of full-lead crystal. The three polar bears are in solid sterling silver. And the combination is dazzling! Each imported sculpture is crafted of hand-gathered crystal and is hand-polished. $135.

Crystal Caper exclusively from The Franklin Mint.

The Franklin Mint
Franklin Center, Pennsylvania 19091

ORDER FORM

Please mail by February 28, 1989.
Limit of one sculpture per person.

Please enter my order for "Crystal Caper" by James Carpenter, to be crafted expressly for me in full-lead crystal and sterling silver.

I need send no money now. Please bill me $33.75* when my sculpture is ready to be sent, and the balance in three equal monthly installments of $33.75* each, after shipment.

**Plus my state sales tax and a total of $3. for shipping and handling.*

Signature ______________________
ALL ORDERS ARE SUBJECT TO ACCEPTANCE

Mr/Mrs/Miss ______________________
PLEASE PRINT CLEARLY

Address ______________________

City ______________________

State/Zip ______________________

85423- 83

LETTERS

To the Editors:

The Shroud of Turin may indeed be a 14th-century fake, but Walter C. McCrone is unlikely to convince very many honest skeptics based on the data reported in "The Shroud of Turin" ["Science and the Citizen," SCIENTIFIC AMERICAN, November, 1988]. However much he may protest to the contrary, his own bias is clearly seen in the ad hominem remark (referring to other scientists) about "big crosses around their necks."

McCrone reportedly took samples of "fibers and other materials lifted from the *surface* of the cloth with sticky tape" (the emphasis is mine). What controls did he use to ensure that he got only original shroud materials and not later accretions? Dust and lint from the 13th century and after could easily have contaminated his sample; a combination of first-century fibers and 15th-century paint might easily yield a 14th-century carbon-14 date.

Earlier reports mentioned a pollen analysis that included pollens unique to the Judean hill country; this finding has not yet been explained away by McCrone. The image on the shroud is a photographic *negative;* McCrone has not offered any evidence that negative imaging was even understood in the 14th century, let alone that it was common practice. How could a 14th-century forger create a new artistic medium *and get it technically correct* by 20th-century standards without any contemporary technology on which to practice his art?

I have no doubt that McCrone has correctly analyzed the coloring as red ocher in a gelatin base, and it may yet be shown that the shroud is a fake. But in his haste to dismiss the whole thing without properly addressing the unanswered questions, it appears that McCrone has overlooked yet a third possibility: that some pious but misguided 14th-century zealot added the dye in an attempt to rescue the image fading from an authentic shroud.

I don't think the Shroud of Turin has been laid to rest quite yet.

TOM PITTMAN

San Jose, Calif.

To the Editors:

I appreciate the opportunity to present further pro-artist arguments. Tom Pittman brings up a number of interesting issues not covered in the short "Science and the Citizen" story. Perhaps I can do a little better here, guided by his comments and questions. I'm sorry I mentioned the crosses worn by STURP (Shroud of Turin Research Project) members, but I thought it inappropriate that STURP officers make such a statement.

My conclusions about the "Shroud" were based on examination of particles *firmly attached* to the more than 43,000 linen fibers picked from the linen surface by the sticky tape. A very thin collagen tempera coating on image-area fibers encapsulated literally billions of submicrometer red-ocher and vermilion pigment particles. They were not loose airborne dust and they were not present in nonimage areas. In a real sense we did have controls: 14 of the 32 tapes were taken from areas of cloth away from the image. I did see and identify many loose extraneous particles on the tapes: minerals, hairs and fibers, air pollution, wax, fungal spores, trichomes (plant hairs), jewelers' rouge, insect parts, starch and so on.

The pollen findings on the "Shroud" are suspect. I found very few pollen grains on tapes from areas where the pollen had been reported. The presence of many more non-European pollens than European—and only pollens from Palestine, Urfa and Turkey—on a cloth with a long history in Europe is implausible.

Negative images are unusual if not unique. But a painting produced to represent a shroud is not a portrait; it shows no light or shadows. Were I painting such an image, I would imagine a cloth over the body and paint the contact areas with artistic gradation, thus darkening the normal highlights. I would be surprised to be told I had created a negative image. The painting of very faint images was common and popular; it is well described in Sir Charles Eastlake's book *Methods and Materials of Painting of the Great Schools and Masters: Volume I.*

My stand on the "Shroud" has been strengthened, although not proved, by the recent carbon date of 1325 ± 65 years. There is now, for example, no possibility that it is an enhanced earlier image, much less chance that the cloth bears pollens from the Near East and a greater chance that the image was painted.

I certainly believe the "Shroud" can confidently be put to rest and classed as a magnificent painting. I hope it happens before I too am put to rest.

WALTER C. MCCRONE

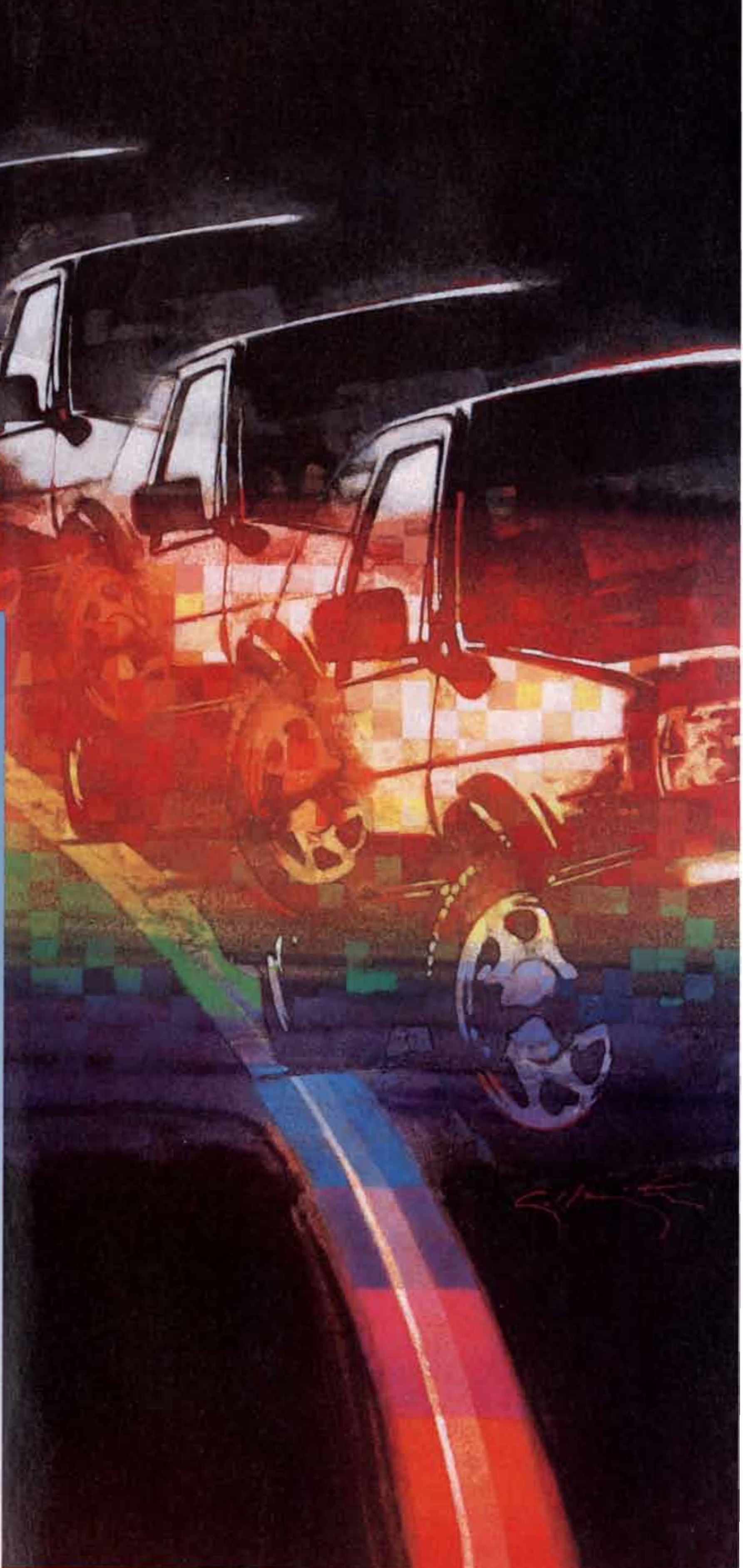
To the Editors:

Jearl Walker's column on the physics of bouncing balls ["The Amateur Scientist," SCIENTIFIC AMERICAN, October, 1988] brought back a memory from my college days in the early 1970's at Trinity University in San Antonio. We had been shown a demonstration of the two-ball drop experiment in a physics class and were very impressed. That afternoon three of us set up our own experiment with a basketball and a softball. We were dropping them, the softball on top of the basketball, from a third-floor walkway between two of the dorms. The basketball was dying on the ground and the softball was shooting up way over our heads. It was going at least 70 or 80 feet into the air.

After we had done this a few times we had attracted quite a crowd, with some onlookers even helping to retrieve the balls and throwing them back up to us. On one drop, however, as Walker warns, the softball shifted to one side. When the two balls hit the ground, the softball flew sideways and went through the resident assistant's window on the ground floor like a cannonball! I am convinced that if the R.A. had been in his room, he would have been severely injured by the ball or the flying glass or would have died of fright on the spot. The broken window cost us $250, in addition to a long lecture from the dorm director.

JOHN C. MCBRYDE

Houston, Tex.



THE BMW 7-SERIES. A NEW GOLDEN AGE OF MOTORING FOR THOSE WHO MISSED THE FIRST ONE.

The big, elegant Bugattis and Duesenbergs from the 20's and 30's combined luxury, performance, and prestige in a way that simply hasn't been seen in the decades since.

So it is a tribute indeed when Automobile Magazine calls a BMW 7-Series the kind of automobile that Bugatti would build if he created "a twenty-first century Royale."

From its flawless body panels to a sculptured engine that mechanics have waited in line just to

THE SANTA BARBARA RACQUET AND POLO CLUB.

look at, this car is crafted to almost otherworldly standards. And that craftsmanship is focused on one end: to provide the most undistracted, enjoyable driving experience available today.

In the hush of its leather interior, your needs anticipated by an unseen staff of computers, you have smooth-riding mastery of "the most complete luxury sedan money can buy" (Automobile).

The cars of motoring's first golden age are in museums now. But your BMW dealer can arrange a test drive of their successor.

THE ULTIMATE DRIVING MACHINE.®

50 AND 100 YEARS AGO

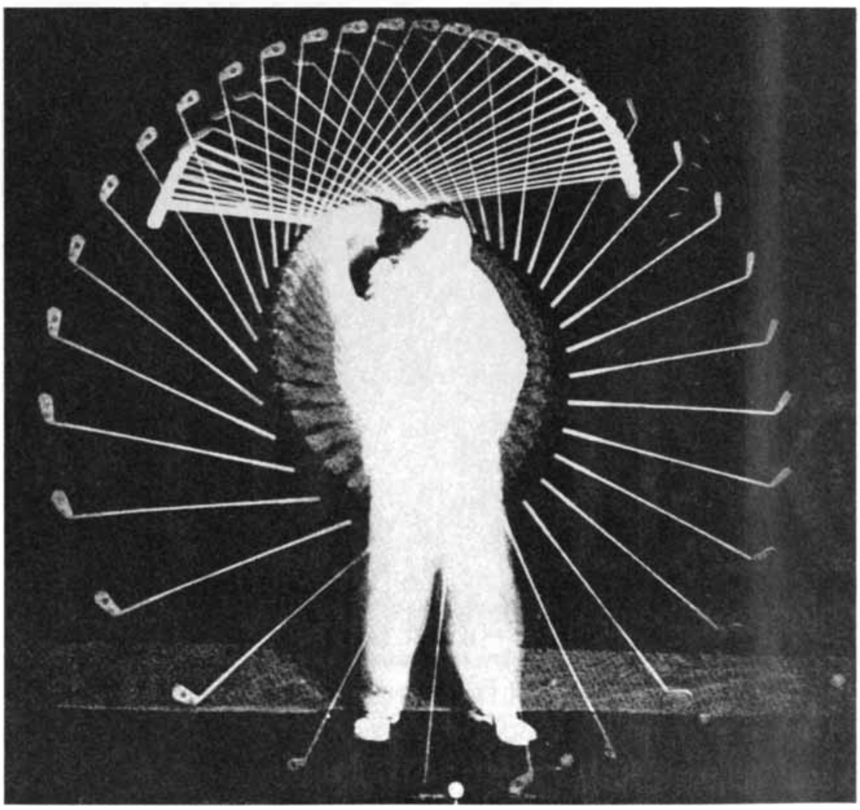
SCIENTIFIC AMERICAN

FEBRUARY, 1939: "A new process for the manufacture of metal objects is now available. It permits the alloying of metals without the customary melting and casting, while entirely new compositions can be produced from such unrelated materials as metals and abrasives. This process of powder metallurgy, which is characterized by the compression of metals in the non-fluid state, is only now coming into its own as a tool of production."

"No clear night is really dark. Only when clouds shut out the sky altogether is there much difficulty in finding one's way in open country. At first thought one would suppose that this faint illumination came from starlight. But photometric measures show there is too much of it to be thus explained. Spectroscopic studies show conclusively that *sodium* atoms are present in the upper regions of the earth's atmosphere in sufficient abundance to give out their characteristic radiation. It is likely the sodium comes from meteorites, which certainly contain sodium and are volatilized, in great numbers, at about this height."

"Something unusual in aluminum safety equipment is the Christiansen safety valve for cows. A rather common complaint of cows is too great indulgence in alfalfa, with the result that the cow becomes badly bloated and in danger of death. The aluminum safety valve, with glass check ball, is permanently inserted in the cow's side. A cow that boasts this latest piece of spare equipment can eat its way through a field of alfalfa and, except for a slight whistle of gas through the valve, suffer no ill effects."

"More and more industrial problems are being solved by the ultra-speed photography process developed by Dr. Harold Edgerton at the Massachusetts Institute of Technology. To the manufacturer of golf balls and clubs, the photograph below, from the newly established research department of A. G. Spalding & Bros., has an important story to tell. In making it, the camera shutter was opened and, as the golfer started his swing, an intense light started flashing, each flash catching the club in a different position."



Harold Edgerton's ultra-speed photography anatomizes a golf swing

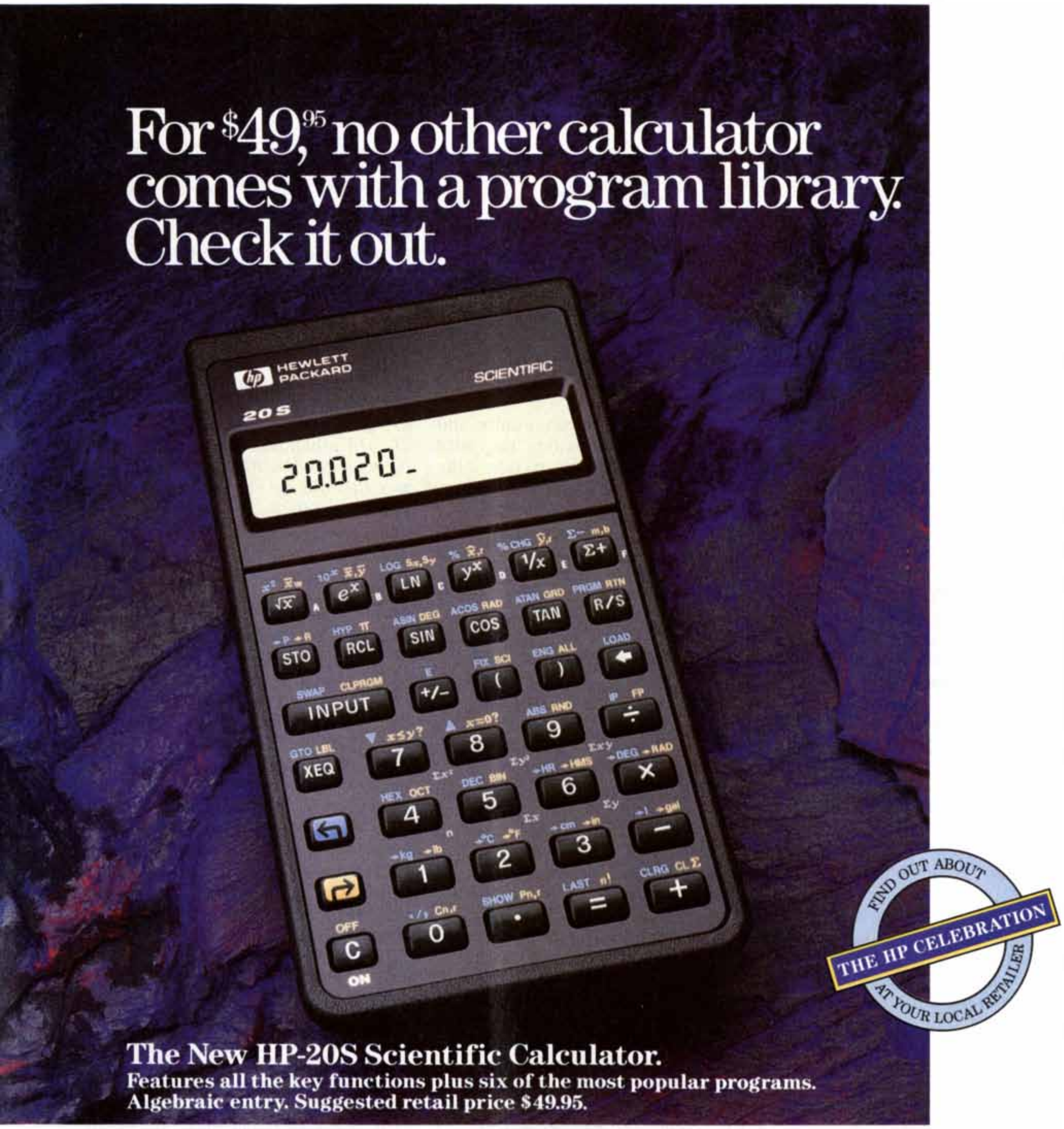
"A portable telephone that gets its power solely from the speaker's voice was announced recently by the Western Electric Company. Its secret of operation rests in a small but powerful magnet. 'Voice currents' are generated when the flux from this magnet is influenced by the motions of a metal diaphragm assembly vibrating under the impact of sound waves. Each telephone is fitted with a small hand crank for generating a signaling tone."

SCIENTIFIC AMERICAN

FEBRUARY, 1889: "Mr. M. Maginn, a mechanical engineer, has designed a device for economically and efficiently developing natural water powers; he suggests its advantageous use at Niagara Falls. The inventor proposes to excavate a cavity or drift under the falls, in front of which the flow of water will be continuous, making a recess in the rock of sufficient depth to accommodate any desired machinery. Within this recess is to be fixed, on permanent foundations, a traveling frame heavy enough to support an overshot steel water wheel of 60 feet diameter, with a main driving shaft and spur gears, with which are to be connected four mammoth dynamos of approximately 2,500 horse power each. The traveling frame is to be moved in and out, so that the wheel shall receive more or less of the force of the current."

"Many of the streets of Astoria, Oregon, are as precipitous as those of our rugged New England towns, and furnish ample grades for the prosecution of that old pastime, sliding down hill. Snow seldom if ever falls, but the climate is so moist that, at the freezing point nights and mornings, a thick coat of white frost covers the planked roadways, which are turned into extempore toboggan slides."

"The Consolidated Gas Light Co. of this city some years ago, in laying a gas main, took advantage of the opportunity to introduce a telephone line in it, suspending it from insulators within the main. Excellent results were attained. On recently opening the main the wire was found to be coated with naphthalene, but the line as such was intact. Such a line is proof against the severest blizzards, and insures communications in all conditions."



Like other scientific calculators in this price range, the new HP-20S offers you all the essential functions you need to solve your problems.

But, unlike other scientific calculators in its class, the HP-20S features a library that includes six of the most-used programs for math, science and engineering.

No need to create and enter programs. We've done it for you. Root finder, numerical integration, complex number operations, 3 x 3 matrix operations, quadratic equation, and curve fitting.

All of which save you the time and effort of creating and entering each program.

And all of the traditional qualities you expect from HP are here. Including a superior processor that makes root finding and integration faster. As well as conditional tests for solving more complex problems.

The New HP-20S Scientific Calculator. Another member of HP's proud new family of quality calculators. See them at your local HP retailer.

HEWLETT PACKARD

SCIENCE AND THE CITIZEN

Econo-SDI

"Limited" antimissile schemes are gaining in popularity

The great champions of the Strategic Defense Initiative—including Secretary of Defense Caspar W. Weinberger, SDI director James A. Abrahamson and of course Ronald Reagan—have retired. SDI budgets are stagnant: four years ago the SDI Organization projected that its annual funding would reach $8 billion in 1989, but last fall Congress provided only $4.1 billion, barely enough to keep pace with inflation. Moreover, all but the most fanatical proponents now acknowledge that Reagan's original vision of making nuclear weapons "impotent and obsolete" is a fantasy.

And yet the SDI is neither dying nor fading away. Indeed, the odds may be better than ever that some form of slimmed-down strategic defense will be deployed in this century. "Reports of the SDI's death," says John E. Pike of the Federation of American Scientists, "have been greatly exaggerated."

Funds for the program, although not increasing at the precipitous rate once hoped for by the Pentagon, have still enabled researchers to produce an array of what Pike calls "real hardware" in the past year or so. This year the SDI Organization plans to test a number of these components, including infrared sensors and particle beams designed to detect enemy missiles and "smart" rockets designed to home in on the missiles and destroy them. If the tests are successful, Pike notes, they may help to quell some doubts about the program's scientific feasibility.

Pentagon officials have also taken steps to make the program more economically palatable. As recently as a year ago, cost estimates for a so-called phase-1 SDI system, which might be capable of shooting down half of the Soviet strategic-missile force, at most, were well over $100 billion. But last fall Abrahamson announced that advances in technology (if not accounting) and a shift of some space-based components to the ground would make it possible to build a phase-1 system for $69 billion. While hardly cheap, this price tag is comparable to that of the B2 ("Stealth") bomber and other large military programs.

There is still considerable skepticism toward the phase-1 plan, not only in the arms-control community and Congress but also among the Joint Chiefs of Staff, who worry that large increases in the SDI will result in cuts in other military programs that they consider more important. Nevertheless, support appears to be growing in Washington for a more limited antimissile system. During last year's campaign George Bush voiced concern that a full-scale strategic defense might be "very expensive" and expressed interest in partial deployment. Bush's national security adviser Brent Scowcroft, who has questioned the feasibility of protecting the entire U.S., recently coauthored a paper that suggested an antimissile system might be needed to protect bomber and submarine bases from missiles launched by Soviet submarines.

Some leading Democrats also seem to have jumped on the limited-defense bandwagon. Early last year Sam Nunn, chairman of the Senate Armed Services Committee, raised the possibility of deploying an "accidental launch protection system" (ALPS) that could shoot down a missile launched in error by the Soviet Union or deliberately by terrorists. More recently Les Aspin, chairman of the House Armed Services Committee, suggested building a system to protect a force of mobile land-based missiles.

Robert S. McNamara, who was secretary of defense from 1961 to 1968 and is now a director of the Arms Control Association, deplores all these schemes. Of the ALPS, he notes that the superpowers could protect themselves from accidental launches much more cheaply and effectively by putting radio-controlled disabling devices on their missiles, and that terrorists are much more likely to smuggle a nuclear bomb into the U.S. on a boat or plane than to launch it on a missile. As for the antimissile defenses proposed by Scowcroft and Aspin, McNamara says there are much cheaper ways to ensure that the U.S. can retaliate to a Soviet first strike—by always keeping some bombers in the air and some submarines at sea, for instance, or through bilateral agreements.

Antimissile defenses, the former defense secretary argues, are inherently destabilizing: they would be much more effective at backing up a first strike than at stopping one. McNamara maintains that this fact, which spurred the U.S. and the Soviet Union to sign the Anti-Ballistic Missile Treaty in 1972, still holds true. By ignoring it, he adds, the U.S. may trigger an arms race that will not enhance U.S. security but imperil it. *—John Horgan*

Stable Orbit

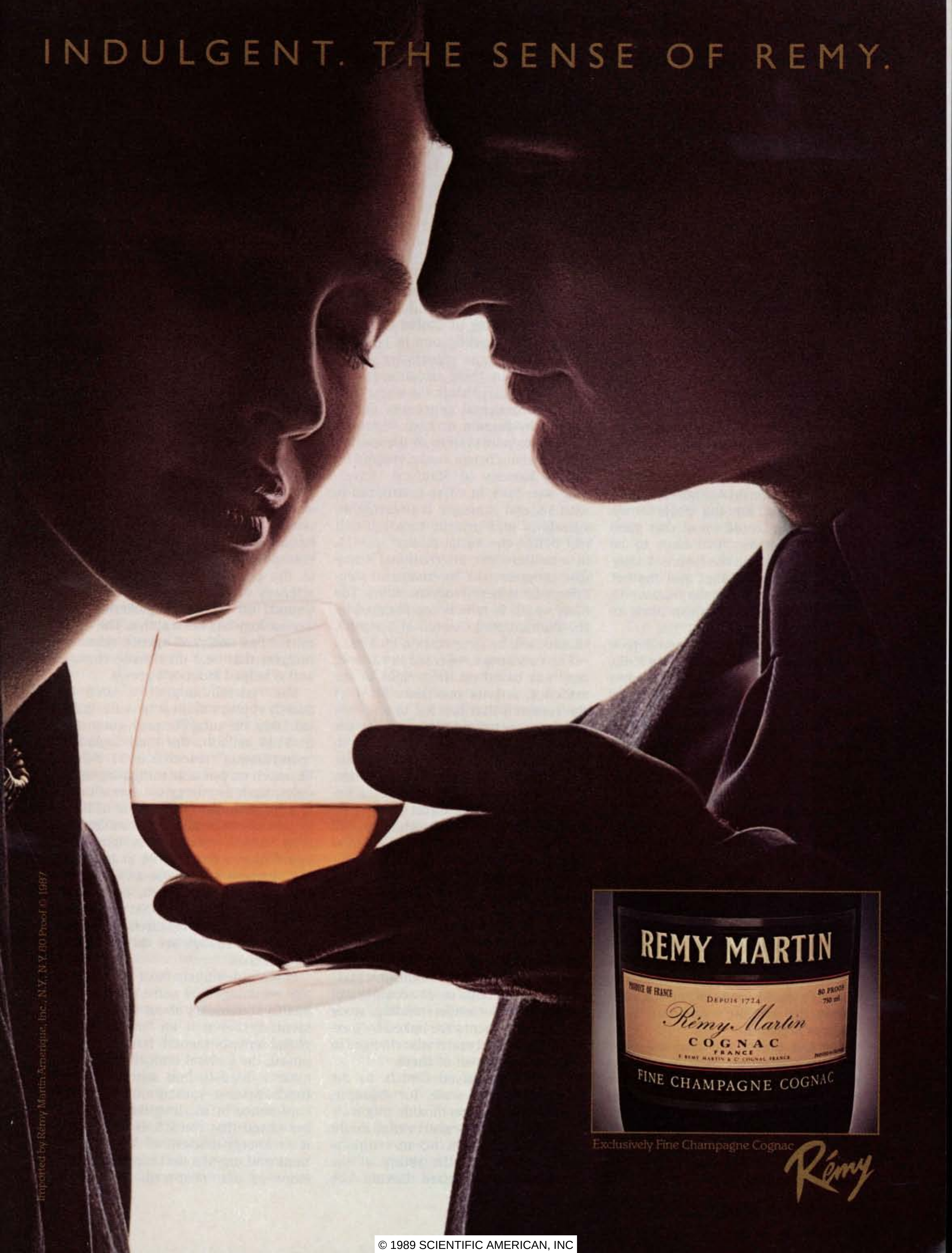
Space reactors cause problems, but a ban is a long way off

Satellites in low orbits tend eventually to tumble into the atmosphere. If they contain nuclear reactors, they can make a dangerous mess. That happened in 1978 when radioactive debris from a Soviet spy satellite scattered over northwestern Canada. Yet commentary by the Soviet news agency Tass suggests that the Soviet Union may be preparing to expand its use of reactors in space. The U.S. also has several orbital reactors under development.

The tip-off about Soviet intentions emerged in a report of a space maneuver executed last October. According to the report, the nuclear-fuel core of a spy satellite, *Cosmos 1900,* was boosted from a decaying orbit into a high "parking" orbit. The report also presented a telling observation from Tass analyst Nikolay Zheleznov. He noted that Soviet specialists had concluded that the success made it "possible to substantially reduce restrictions on the scope of application of nuclear power plants in outer space."

The Soviet Union uses low-power reactors for its RORSAT spy satellites, which carry radars that track and target NATO ships. The U.S. does not now use reactors in satellites, although the Department of Energy is developing five powerful models for possible use in a "Star Wars" system.

Tass's analysis is unwelcome news


INDULGENT. THE SENSE OF REMY.
REMY MARTIN
PRODUCE OF FRANCE
DEPUIS 1724
80 PROOF
750 ml
Rémy Martin
COGNAC
FRANCE
FINE CHAMPAGNE COGNAC
Exclusively Fine Champagne Cognac
Rémy
Imported by Remy Martin Amerique, Inc., N.Y., N.Y. 80 Proof © 1987

for opponents of orbiting reactors. At least two groups have argued for a ban on reactors in near-earth orbit. They are the Federation of American Scientists (FAS) and the Committee of Soviet Scientists against the Nuclear Threat (CSS). Both have drawn attention to the severe radioactive contamination that could result if a large reactor burned up in the atmosphere. Roald Z. Sagdeev, chairman of the CSS and until recently director of the Soviet Institute of Space Research, has written to Senator J. Bennett Johnston that a worst-case accident could "be comparable to the long-term consequences of Chernobyl."

Even in orbit the devices create problems. Gamma rays generated by such reactors concern engineers working on the *Gamma Ray Observatory*, an earth-orbiting research platform scheduled to be launched in 1990. Gerald H. Share of the Naval Research Laboratory calls gamma-ray pollution "a very serious problem." According to Donald A. Kniffen, NASA's chief scientist for the observatory, more reactors could mean that gamma-ray astronomy "will have to be done differently" in the future. A Japanese observatory, *Ginga*, and another planned NASA mission, the *Nuclear Astrophysics Explorer*, are also likely to be affected.

The *Solar Maximum Mission* detected gamma rays from RORSAT's in 1980, a fact that was classified until a few months ago. (It has been speculated that the delay was occasioned by a quiet analysis intended to reveal details about the Soviet reactors.) Orbiting reactors produce gamma rays not only directly (at close quarters) but also by means of positron emission. Positrons from a reactor move along the earth's magnetic field lines for thousands of miles, generating gamma rays if they strike a satellite.

Kniffen is currently devising ways to protect the *Gamma Ray Observatory's* sensors. The worst-affected instrument is one designed to start recording data when it detects natural gamma-ray "bursts" (which are of unknown origin). False alarms caused by RORSAT's could waste observing time unless the satellite's on-board computer is prepared.

The FAS points out that the gamma rays emitted by orbiting reactors mean that a ban would be verifiable. Moreover, an FAS delegation was told by a senior Soviet foreign-ministry official that the Soviet Union would "very seriously consider" a ban on orbiting nuclear reactors if one was proposed by the U.S. In view of the U.S.'s large investment in nuclear power systems intended for use in space, that seems unlikely in the near future. *—Tim Beardsley*

Future Imperfect

Concern about global climate percolates up to government

"The problem of global environmental change," the National Research Council declares, "is crucial and urgent." That assessment is made in a major new study, "Toward an Understanding of Global Change," scheduled for publication in January. The council, whose reports are generally noted for their circumspection, observes bleakly that in the worst case the environmental processes set in motion by human activity "threaten the life-support system of the earth."

The global-change study, chaired by Harold Mooney of Stanford University, was done in close collaboration with Federal agencies. It presents an agenda of high-priority research that will define the initial role of the U.S. in a multidecade international scientific program, the International Geosphere-Biosphere Program (IGBP). The IGBP, which is now being planned by the International Council of Scientific Unions, will be in operation by 1992.

The consensus reflected in the NRC report is based on the weight of the evidence, not its precision. Most of the research that has led to the now widely accepted predictions that the global climate will change significantly within decades has relied on computer models of atmospheric change. Two principal concerns emerge: an enhanced greenhouse effect due to accumulating trace gases, which would warm the earth's surface, and a depletion of the stratospheric ozone layer by chlorofluorocarbons (CFC's), which is expected to increase levels of ultraviolet light.

Such models rely on simple assumptions about how the ocean, the land's surface and living things will react to the changing conditions. Making the predictions more accurate will require a better understanding, since all the components are linked by fluxes of energy and materials: changes in one could affect all of them.

If warming caused forests to die back on a large scale, for example, atmospheric carbon dioxide might increase more rapidly and rainfall might decline. Changes in ocean temperature would affect the ability of the ocean to absorb carbon dioxide. The council therefore recommends detailed studies of biogeochemical dynamics: how critical chemicals move between the air, the oceans and the land. It also wants to see more attention paid to how ecological systems respond to changes in conditions. For instance, a modest increase in ultraviolet light might disrupt the reproduction of plankton that form the base of the marine food chain.

The council called for continued inquiry into how human activity affects the environment and vice versa. It emphasized the importance of studies of the earth's climatic history, geophysical processes and the atmosphere's responses to changes in the sun's radiating energy.

Governments have begun to think about global change. An Intergovernmental Panel on Climate Change, convened at the request of the World Meteorological Organization and the United Nations Environment Programme, met in November in Geneva and set up working groups to review scientific understanding, socioeconomic impacts and strategies. There has already been some discussion of a convention to protect the atmosphere. In the U.S. the committee on earth sciences of the Federal Coordinating Council on Science, Engineering and Technology has been active. The committee has analyzed agency research budgets that bear on climate change, and it helped Mooney's group.

The Federal budget for such research appears modest in scale. In fiscal 1989 the total for such programs is $134 million; the total including "contributing" research is $1 billion. Research on possible mitigating strategies, such as energy conservation, is at a low ebb: the Department of Energy will spend a mere $166 million on it in 1989. Yet satellite systems that could play a crucial role in monitoring changes, such as NASA's planned Earth Observing System, could cost hundreds of millions. Several existing international research programs on climate change are threatened by budget cuts.

Environmentalists have been generally encouraged by some of President Bush's statements about the environment. Yet even if an "initiative" on global environmental issues is proposed, the Federal budget deficit will make it hard to find significant new funds. Several speakers at a climate conference in Washington in December noted that the U.S. is one of the least energy-efficient of Western nations and urged a tax on carbon emissions as one near-term strategy to alleviate greenhouse warming. That would be politically difficult: two of Bush's most winning campaign pledges, "I am an environmentalist" and "No new taxes," may come together to haunt him. —*T.M.B.*

PHYSICAL SCIENCES

In Memoriam
The 300-foot radio telescope has collapsed

As Route 28 snakes northward through Deer Creek Valley in West Virginia, it carries you past settlements too small to be incorporated, dilapidated farms, houses built from trailers, traffic signs riddled with bullet holes. Trees, streams, more trees. Then suddenly, just south of Green Bank, the road crests and bends, and staring directly at you from the valley floor is the giant dish of the 300-foot radio telescope. It is one of the world's extraordinary sights. Or it was until November 15 at 9:45 P.M., when the great telescope collapsed.

Drivers from Cass and Dunmore were in a state of shock the next morning when they rounded the bend to find that this monumental fixture of the landscape for 25 years had simply vanished. I can imagine their explanations: when the National Radio Astronomy Observatory moved to Green Bank way back then, folks blamed the telescopes for ruining both the TV reception ("The telescopes suck in the signals") and the winter ("It used to snow more").

Perhaps. But the NRAO also brought jobs, tourists and students. When I was a summer research assistant at Green Bank in 1974 and 1976, my adviser used to tell how the 300-foot was "thrown together" because construction on the 140-foot—a much more complicated instrument—was running behind schedule. They finished the 300-foot in 1962 at a cost of about $850,000, compared with $14 million for the 140-foot.

The primary cost-cutting measure lay in designing the 300-foot so that it could be steered only in a north-south direction; it relied on the rotation of the earth to carry cosmic radio sources through its beam. Then too, its surface was not solid but a mesh, which limited its operations to lower frequencies than those attainable by some of the other telescopes.

The designers often joked that they expected the 300-foot to collapse after five years. Yet the instrument, "thrown together" out of gossamer, outlasted all expectations. Serviced weekly, for 26 years it surveyed the skies for new radio sources and took part in a few searches for extraterrestrial intelligence.

"Thrown together" in 1962 for less than $1 million, the telescope performed beyond all expectations

300-FOOT ANTENNA of the National Radio Astronomy Observatory in Green Bank, W.Va., helped to establish that pulsars are rapidly rotating neutron stars.

The 1967 detection of a pulsar in the Crab nebula was the 300-foot's most famous achievement. Pulsars had been discovered a few months earlier, but their physical nature remained unknown. Only with the discovery of the Crab pulsar, which was spinning at a rate of 30 times per second, did theorists realize that pulsars must be rapidly rotating neutron stars and not white dwarfs or "little green men."

The 300-foot did have its problems. Worried that the accumulation of snow on the dish might cause it to collapse, members of the staff brought in a jet engine to blast the snow away. Fortunately they never used it. Birds were another difficulty: they liked to roost in the receiver horns, far above the dish. To avoid climbing up to the antenna focus all the time to clean out the nests, somebody installed an electronic scarecrow, a tone generator that produced cosmic-sounding bleeps and bloops that were audible all around the area. We would tell tourists that the bleeps were signals being received from space—"music of the spheres"—and the visitors went away satisfied.

For the past five years or so the observatory's interferometer, consisting of three 85-foot dishes, has been run by the Navy for nonastronomical research. And so when the 300-foot collapsed in November, the future of world radio astronomy looked bleak indeed. But Senator Robert C. Byrd of West Virginia has come out strongly in favor of constructing a new one, and a special bill is expected before Congress. NRAO scientists are optimistic that a 100-meter telescope will be constructed on the site of the 300-foot. May it prove to be a worthy successor. —*Tony Rothman*

Hidden Chemistry
Does the earth's core react with the mantle?

The earth's boundaries are eventful places. Life evolved where the solid earth meets the atmosphere; mountain building, volcanism and earthquakes take place at the junctions of the plates making up the earth's surface. Some 2,900 kilometers down, the solid rock of the earth's mantle adjoins the liquid iron alloy of the outer core. Chemically this may be the most active boundary of all, according to results reported at the winter meeting of the American Geophysical Union.

The dynamic nature of the region was known from earlier studies, based on earthquake waves that penetrate the earth and on measurements of its magnetic field. Just above the core-mantle boundary lies a 200-kilometer-thick layer from which seismic waves scatter unpredictably. The magnetic field, which is generated near the boundary by the flowing metal of the outer core, has stationary regions of anomalously high and low intensity as well as a general westward drift.

At the AGU meeting investigators proposed that the core-mantle boundary is the site of intense chemical activity as well. Raymond Jeanloz and Xiaoyuan Li of the University of California at Berkeley and Elise Knittle and Quentin Williams of the Santa Cruz campus base their proposal on laboratory simulations that rely on a device called a diamond-anvil cell. The cell duplicates the enormous pressures of the deep earth by multiplying the force of a simple screw and transmitting it to a tiny sample of minerals trapped between the faces of two diamonds. A laser shining through the transparent diamond heats the sample to mantle temperatures.

Earlier research in diamond-anvil cells had shown that the high pressures of the lower mantle transform its components to a mineral known as silicate perovskite, a magnesium/iron silicate, or $(Mg,Fe)SiO_3$. The studies also yielded a plausible temperature for the core-mantle boundary. It must lie above the melting point of the core's iron alloy at such pressures (since the outer core is liquid) but below the melting point of the perovskite. By determining the melting temperatures of both materials at the pressures of the core-mantle boundary, Jeanloz and his colleagues were able to assign the boundary a temperature of between 4,000 and 6,000 degrees Celsius, much higher than earlier estimates.

It turns out that the perovskite and iron react fiercely under those conditions of temperature and pressure. To create a microcosm of the boundary, the workers embedded bits of iron foil in the perovskite, compressed the sample and heated it to the boundary temperature—hot enough to melt the iron but not the mineral. Then they opened the cell and analyzed the contents. Where metal met rock a broad reaction zone had developed, in which regions of the minerals stishovite (SiO_2) and iron-depleted silicate perovskite ($MgSiO_3$) alternated with pockets of iron/oxygen and iron/silicon alloy. Significantly, stishovite is stiff, which speeds the propagation of seismic waves, and electrically insulating. The alloys are softer—hence seismically slow—and highly conductive.

Extrapolating the results to the core-mantle boundary, the investigators envision a reaction zone that might encompass the entire 200-kilometer thickness of the scattering layer. The juxtaposition of stishovite with pockets of alloy would yield sharp variations in seismic-wave velocity that could account for the seismic scattering. And the million- or billionfold lateral variations in electrical conductivity could sculpture the earth's magnetic field.

Most workers have assumed that the conductivity of the lower mantle is negligible, so that the magnetic field generated in the outer core reaches the surface without much distortion. Investigators have tended to attribute observed anomalies in the field to turbulent flow in the core. Pockets of conductive alloy in the overlying mantle, however, could also distort the field: a conductive blob would impede field lines being swept along by the flowing metal, causing them to pile up. At the surface the magnetic flux would be reduced in one region and intensified in an adjacent one.

Jeremy Bloxham of Harvard University, an investigator of the earth's magnetism, is not convinced that conductivity variations in the lower mantle will turn out to be a broad explanation for anomalies in the magnetic field. He does think a region of high conductivity under the Pacific could account for the large stationary zone of low magnetic flux seen there. Jeanloz points out that the core-mantle boundary under much of the Pacific is seismically slow, as one would expect of a large region of alloy produced by core-mantle chemistry. *—Tim Appenzeller*

Blame It on the Moon

Australian "solar varves" turn out to be mostly lunar

The Elatina formation, an outcropping of beige sandstone in southern Australia, is composed of sediments laid down during the Precambrian era, nearly 700 million years ago. Measurements of the thickness of successive sediments show that they wax and then wane every dozen or so layers. Over what time scales were the sediments deposited, and what explains their periodic change in width?

In 1983 George E. Williams of the Broken Hill Proprietary Company Ltd. in Adelaide thought he had the answer. He proposed that the Elatina sediments had been laid down in annual increments, probably at the bottom of a lake bed fed by melting glaciers. The fluctuations in their thickness, he said, reflected variations in solar radiance that accompany the solar cycle, a period from 10 to 14 years long during which sunspots increase and decrease in number [see "The Solar Cycle in Precambrian Time," by George E. Williams; SCIENTIFIC AMERICAN, August, 1986].

Williams has overturned his own hypothesis. After analyzing Precambrian sediments from another part of southern Australia, he recently concluded that both they and the Elatina sediments were shaped primarily by the moon's influence, not the sun's. Charles P. Sonett, Stephen A. Finney and Cameron R. Williams of the University of Arizona, who have also studied the Elatina sediments, have seconded the new interpretation.

Writing in *Nature*, the Arizona team proposes that the sediments were deposited not in a lake but in a lagoon or bay and that each layer was generated not over the course of a year but during a single tide. The dozen-layer period apparently marked the completion not of a multiyear solar cycle but of a semimonthly lunar cycle, during which tides wax and wane in strength.

This finding marks a setback for investigators seeking a connection between periodic changes in the sun's behavior and the earth's climate. But it has provided insights into how the Precambrian era differed from our own. Based on an extensive analysis of the new data, Sonett and his colleagues estimate that the moon was about 8,200 miles closer to the earth than it is now (today's mean distance is 238,700 miles) and revolved around the earth faster (14 rather than 12 times a year). The earth also seems to have rotated about 400 times a year rather than 365. *—J.H.*

First Impressions

Explosives help a sculptor to make bas-reliefs in metal

Salvador Dali once exploded a bomb filled with nails against a copper plate, producing a striking but random pattern. Many other

PULSE-QUICKENING COMFORT.

INTRODUCING A COMPLETELY DIFFERENT MERCURY COUGAR.

We did more than give Cougar a sleek, new exterior. Much more. We gave it a longer wheelbase for a more comfortable ride. And an all-new suspension design with a lower, wider stance for greater stability and control. Plus new, speed-sensitive variable-assist power steering for improved road "feel" and control at highway speeds. Inside, it was obviously designed by people who know that cars are for driving. Experience the comfort and control of a Mercury Cougar at your Lincoln-Mercury dealer today. For more Cougar information, call 1-800-822-9292.

MERCURY

WHERE COMFORT AND CONTROL ARE ONE.

LINCOLN-MERCURY DIVISION Ford Buckle up—together we can save lives.

artists have also utilized explosives in their work, but the results have generally been unpredictable. Now Evelyn Rosenberg, a sculptor working at the New Mexico Institute of Mining and Technology, has harnessed the energy of explosives to create predetermined bas-reliefs in brass, copper and stainless steel.

Rosenberg's first attempts were disappointing, she wrote recently in *Leonardo,* a journal of technology in the arts. Sometimes the explosives blew holes through the metal or flung the materials into the air. After 80 trials she was finally able to manipulate such variables as the thickness of the metal and the energy of the explosion to create sculptures, which she calls detonographs. Rosenberg has now produced 70 detonographs, seven of which hang above the entrance to the New Mexico Museum of Natural History in Albuquerque.

The power of plastic explosives can be harnessed to sculpt metal bas-reliefs called detonographs

EXPLOSIVES fused and molded stainless steel, brass and copper to form this four-foot-wide bas-relief titled* Forces and Signs. *Evelyn Rosenberg created it for Gaitersburg Junior High School in Maryland. The photograph is by Jerry Goffe.

To create a detonograph, Rosenberg first sculpts a bas-relief in plaster to form a mold. She covers the mold with a three-by-four-foot panel of brass, copper or stainless steel. On the metal panel she sometimes places thin colored metallic foils and such objects as cloth, string or leaves. At a blast site in Socorro, N.Mex., a technician covers the layers with Du Pont's C-1 Detasheet, a plastic explosive composed of pentaerythritol tetranitrate, cellulose nitrate and a plastic binder.

When the technician detonates this explosive sandwich (generating a 20-foot fireball), the explosion drives the metal panel into the mold and thereby reproduces the bas-relief in metal in the instant before the mold shatters. The tremendous pressure of the explosion clads the panel with the metal foils, which provide colored accents; the cloth, string or leaves are driven into the panel at 6,800 meters per second and provide texture by leaving their imprint before disintegrating in the flames. To complete the detonograph Rosenberg cleans and polishes the panel.

Although works similar to detonographs are made by die-stamping, casting at a foundry or repoussage (in which soft metal sheets are shaped by hammering), Rosenberg maintains that her method is more suitable for individual works of art, is less time-consuming and is relatively inexpensive. Her method also makes it possible to clad one kind of metal with another, something that cannot be accomplished employing the other techniques. She comments: "My expression as a sculptor is reinforced by the magic, primitive quality of explosive art." *—Russell Ruthen*

BIOLOGICAL SCIENCES

POU! Goes the Homeobox

Developmental DNA sequences are found in puzzling places

The homeobox, once thought to be the holy grail of developmental biology, has recently become a trickier cup to drink from. The homeobox is a short segment of DNA found in genes that appear to control key events in the embryonic develop-

Tandy® Computers: The broadest line of PCs in America.

The Tandy 1000 TL

286 power with MS-DOS® and DeskMate® built in.

The Tandy 1000 TL is a powerful computer for personal and business use. Its 80286 microprocessor gives you extraordinary speed and processing power. Plus, the 1000 TL comes with MS-DOS and the DeskMate Graphical User Interface built in, so you can be up and running in seconds, using plain-English commands.

DeskMate features ten applications that let you write reports and letters, prepare budgets, file, draw colorful pictures, create and play back songs and more. Plus, there's PC-Link®, an online information service.

You also get the latest in computer-audio technology. When you use DeskMate's sound editor, you can record and edit voice, music or any analog source onto diskettes.

The 640K Tandy 1000 TL comes with a 3½″ disk drive and has room for an additional 3½″ and 5¼″ drive. A parallel printer adapter, RS-232 serial port, two joystick ports, a clock/calendar and five expansion slots are all standard. You also get a 101-key enhanced keyboard for the ultimate combination of power, ease of use and affordability.

The new generation Tandy 1000 TL. From the best-selling family of PC Compatibles made in America.

Send me a 1989 RSC-20 Computer Catalog

Mail to: Radio Shack, Dept. 89-A-921
300 One Tandy Center, Fort Worth, TX 76102

Name ______________________

Address ______________________

City ______________________

State ____________ ZIP ____________

Phone ______________________

Tandy Computers: Because there is no better value™.

MS-DOS/Reg. TM Microsoft Corp. PC-Link/Reg. TM Quantum Computer Services.

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

ment of a very wide range of species. It has been assumed that homeobox proteins (the proteins that are encoded by such genes) function by binding to DNA and thereby controlling gene expression, but hard evidence was lacking.

Now several groups report conclusive evidence that homeobox proteins do bind to DNA and regulate the transcription of RNA molecules (the first step in gene expression). Yet the new findings were made in an unexpected context: not in the differentiating cells of an embryo but in the mature tissues of several mammalian species. It appears that some homeobox proteins determine the fate of mature cells rather than embryonic ones. Thus while the mechanism of the homeobox has been put on a sounder footing, its overall role has been considerably complicated.

The homeobox was first identified in 1983 as a common region in several genes that control the division of the embryo of the fruit fly *Drosophila* into segments and help to determine the segments' fate (which segment becomes a leg as opposed to becoming a wing, for example). Since that discovery many additional homeobox-containing genes have been found in *Drosophila,* and similar DNA segments have been implicated in the embryonic development of worms, frogs and mice. In all these cases genes containing the homeobox seem to serve as "master" genetic elements, interacting with a cascade of other genes to send a cell down a specific pathway of development.

The three mammalian genes that have recently been shown to include the homeobox act in a different context. Each of these genes encodes a "transcription factor," which, by binding to a specific stretch of DNA, influences the rate at which messenger RNA molecules are transcribed. The transcription factors are known as OCT-1, OCT-2 and Pit-1, and each acts in a characteristic range of tissues. At least two of the transcription factors—OCT-2 and Pit-1—serve to determine the final identity of a particular set of mature tissues.

OCT-2 is made in immune-system cells known as *B* cells; it causes them to produce large quantities of antibodies as they develop into the mature type called the plasma cell. The transcription factor has its effects by triggering immunoglobulin genes, which encode the components of antibodies. It had been known that OCT-2 acts by binding to DNA. In a recent paper in *Nature* a group led by Robert G. Roeder of Rockefeller University reports that they have cloned the gene for OCT-2 and shown that it contains a homeobox. This result, according to the investigators, is "surprising but significant."

Pit-1, on the other hand, has its effects in the pituitary, where it apparently causes two closely related cell populations to make their characteristic hormones: growth hormone and prolactin (which sustains lactation). Now two groups at the University of California at San Diego School of Medicine (one led by Michael Karin, the other by Michael G. Rosenfeld) report in *Cell* that they have cloned the Pit-1 gene. Both groups agree that the gene for Pit-1 includes a homeobox.

The discovery that the homeobox functions in mature cells was not the end of the surprise. Flanking the homeobox in each of the three newly characterized genes was another common segment of DNA. The homeobox contains 180 nucleotides coding for 60 amino acids; the new region is slightly larger, including enough DNA to encode 75 amino acids. Because it was discovered in the genes for Pit and OCT and also in the roundworm *Caenorhabditis elegans* in a gene called *unc-86* (which has a similar role in specifying the type of mature tissues), the newly discovered conserved region has been named POU (pronounced "pow").

Although the discovery of the combination of POU and the homeobox is intriguing, it clearly raises as many questions as it answers. Among the more significant puzzles are: What is the common thread between the effects of such genes in early development and in the terminal differentiation of mature tissues? What are the evolutionary relations between these two different functions? Of even greater immediate interest is the question of how the homeobox-POU combination works at the level of molecular detail.

The simple answer to the last question would seem to be that each homeobox protein binds to a unique DNA sequence, thereby triggering a unique gene or combination of genes. Unfortunately the simple answer appears to be false. Unlike OCT-2 and Pit-1, the OCT-1 protein is "ubiquitous," or present in not one but many cell types; it triggers the release of proteins synthesized in many different types of cell. Furthermore, the type-specifying protein OCT-2 and the ubiquitous protein OCT-1 bind to the same DNA sequence. A simple model in which the identity of a cell results from a unique match between a homeobox protein and a particular piece of DNA no longer holds. If the new results are confirmed, then it would seem that the grail holds a beverage of a larger quantity and a lesser potency than once was thought. *—John Benditt*

A Breed Apart

Finicky flies lend credence to a theory of speciation

What is the origin of species? Conventional wisdom holds that a new species can arise when members of the same species are separated by geographic barriers; the barriers prevent interbreeding between the two populations, and heritable differences are established as the populations adapt to their different environments. Since Darwin's time, however, it has been proposed that species can also originate "sympatrically"; that is, some members of a population may become reproductively isolated from the rest of the population when no physical barriers to mating exist. Now entomologists may have found the first genetic evidence that sympatric speciation does in fact take place.

The evidence comes from studies of a fruit fly called *Rhagoletis pomonella.* The insect is a parasite of the hawthorn tree and its fruit, which is commonly called the thorn apple. The life cycle of the hawthorn fly revolves around the thorn apple: the fruit serves as a rendezvous for courtship and mating, the females lay their eggs in it and the larvae feed on it.

At least some of them do. About 150 years ago *R. pomonella* began to infest apple trees as well as hawthorns. Fly populations diverged into two so-called host races: one that prefers apples and one that prefers hawthorns. A former classmate of Darwin's named Benjamin Walsh first recognized the new apple-fly race and suggested that the differential host preferences could lead to speciation.

In the early 1960's a Harvard University graduate student named Guy L. Bush set out to discredit Walsh's theory by demonstrating that apple and hawthorn flies were not becoming two different species. But as Bush learned more about *Rhagoletis* he began to doubt his own premise. Generally organisms are considered separate species if they do not interbreed. Bush realized that if the host fruit provides the site for *Rhagoletis* mating, breeding between apple and hawthorn flies

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

Speed Learning has replaced speed reading. It's a whole new way to read and learn. It's easy to learn . . . lasts a lifetime . . . applies to everything you read. It may be the most productive course you've ever taken.

Do you have too much to read and too little time to read it? Do you mentally pronounce each word as you read? Do you frequently have to go back and reread words or whole paragraphs you just finished reading? Do you have trouble concentrating? Do you quickly forget most of what you read?

If you answer "yes" to any of these questions—then here at last is the practical help you've been waiting for. Whether you read for business or pleasure, school or college, you will build exceptional skills from this major breakthrough in effective reading, created by Dr. Russell Stauffer at the University of Delaware.

Not just "speed reading"—but speed reading-thinking-understanding-remembering-and-learning

The new *Speed Learning Program* shows you step-by-proven -step how to increase your reading skill and speed, so you understand more, remember more and use more of everything you read. The typical remark from over one million people taking the Speed Learning program is, "Why didn't someone teach me this a long time ago." They were no longer held back by the lack of skills and poor reading habits. They could read almost as fast as they could think.

What makes Speed Learning so successful?

The new *Speed Learning Program* does not offer you a rehash of the usual eye-exercises, timing devices, costly gadgets you've probably heard about in connection with speed reading courses or even tried and found ineffective.

In just a few spare minutes a day of easy reading and exciting listening, you discover an entirely new way to read and think—a radical departure from anything you have ever seen or heard about. Speed Learning is the largest selling self-study reading program in the world. Successful with Fortune 500 corporations, colleges, government agencies and accredited by 18 professional societies. Research shows that reading is 95% *thinking* and only 5% eye movement. Yet most of today's speed reading programs spend their time teaching you rapid eye movement (5% of the problem) and ignore the most important part (95%) *thinking*. In brief, *Speed Learning gives you what speed reading can't.*

Imagine the new freedom you'll have when you learn how to dash through all types of reading material *at least* twice as fast as you do now, and with greater comprehension. Think of being able to get on top of the avalanche of newspapers, magazines and correspondence you have to read . . . finishing a stimulating book and retaining facts and details more clearly and with greater accuracy than ever before.

Listen-and-learn at your own pace

This is a practical, easy-to-learn program that will work for you—no matter how slow a reader you think you are now. The *Speed Learning Program* is scientifically planned to get you started quickly . . . to help you in spare minutes a day. It brings you a "teacher-on-cassettes" who guides you, instructs, encourages you, explaining material as you read. Interesting items taken from *Time* Magazine, *Business Week, Wall Street Journal, Family Circle, N.Y. Times* and many others, make the program stimulating, easy and fun . . . and so much more effective.

Executivies, students, professional people, men and women in all walks of life from 15 to 70 have benefited from this program. Speed Learning is a fully accredited course . . . costing only 1/5 the price of less effective speed reading classroom courses. Now you can examine the same, easy, practical and proven methods at home . . . in spare time . . . without risking a penny.

Examine Speed Learning FREE for 15 days

You will be thrilled at how quickly this program will begin to develop new thinking and reading skills. After listening to just one cassette and reading the preface you will quickly see how you can achieve increases in both the speed at which you read and in the amount you understand and remember.

You must be delighted with what you see or you pay nothing. Examine this remarkable program for 15 days. If, at the end of that time you are not convinced that you would like to master Speed Learning, simply return the program and owe nothing. See the coupon for low price and convenient credit terms.

No-Risk Guarantee

If, after reviewing *Speed Learning* for 15 days, you are not delighted in every way, you may return the materials in the condition you received them *and owe nothing.*

BONUS GIFT

Order *Speed Learning* today and get this Deluxe, Compact Stereo Cassette Player . . . *at no additional cost.* It's perfect for listening to your *Speed Learning* tapes, anywhere, anytime.

ORDER FORM

(Send No Money Now)

learn INCORPORATED Dept. DZ-24
113 Gaither Drive, Mt. Laurel, NJ 08054-9987

☐ **YES!** I want to increase my earning power. Please rush me the complete **Speed Learning** program and, later, bill me in four equal payments of $24.95*. Also, send me my Compact Stereo Cassette Player, at no additional cost. **My satisfaction is guaranteed.**

Name ______________________

Address ______________________

City ______________ State ________ Zip ________

Phone No. (   ) ______________________

Signature ______________________

*Plus $6 shipping and handling. For New Jersey residents sales tax will be added.

DISCRIMINATING **Rhagoletis pomonella** ***flies feed and breed on either apples or thorn apples—but not both.***

must be minimized. By the time he got his Ph.D., Bush had become convinced that the flies were speciating sympatrically.

Genetic analyses done by Bush and Jeffrey L. Feder of Michigan State University with Charles A. Chilcote, now at the University of Michigan, support his conviction. The studies, published recently in *Nature*, reveal significant differences in the genetic makeup of the two races—differences that could not be maintained if the flies mated randomly. A report from Bruce A. McPheron and his colleagues at the University of Illinois at Urbana-Champaign confirms those findings. And D. Courtney Smith, now at Pennsylvania State University, has shown that the timing of the adult's emergence from the pupa, which is synchronized with the ripening of either the apple or the hawthorn fruit, is an inherited trait.

It seems, then, that the two populations of flies are on the way to becoming separate species. While noting that "allopatric" speciation—the kind imposed by geographic separation—does explain the origins of most larger animals, Bush says a substantial percentage of the 600,000 species of parasitic insects may be the product of sympatric speciation. Without a historical record, however, it is hard to dismiss the allopatric alternative.

The appeal of *R. pomonella*'s case is that the history of its host races is known. Although there might still be some interbreeding between apple and hawthorn flies, Bush thinks it is unlikely that the two populations will fuse again. How long will it be, then, until the apple fly can be called a species in its own right? "Hard to say," Bush's colleague Feder says. "Check back with me in a few thousand years." —*Karen Wright*

MEDICINE

Exercising Choice

Case (almost) closed: fitness does seem to prolong life

Several classic studies have indicated that people who exercise are less likely to die from cardiovascular disease (CVD, which usually strikes as a heart attack) than are their more sluggish counterparts. Most physicians accept the fact that some exercise is beneficial, but it has not been clear whether the lower rate of CVD in active people is a direct benefit or a secondary effect—a consequence of fitter people weighing less and having lower blood pressure and better cholesterol profiles. A few skeptics have even argued that fit people are healthy not because they exercise but because of hidden factors such as favorable genes that also make them inclined to be active.

Now a study has been published that puts the case for a direct benefit on much firmer ground. Although final proof is elusive—it would require prescribing exercise programs lasting for decades—the study shows that physical fitness itself seems to be beneficial. Indeed, two of the senior investigators now maintain that inactivity should be recognized as a primary risk factor for CVD.

In the *New England Journal of Medicine* Lars-Göran Ekelund of the University of North Carolina and his colleagues report that they studied 3,106 adult male Americans who showed no clinical evidence of CVD. At the outset the workers obtained an objective measurement of the subjects' physical fitness by measuring performance on a treadmill, an improvement over studies that rely on self-reported exercise. Then they recorded deaths from CVD over the next eight years.

The number of men in the least-fit fourth of the sample who died from CVD was 8.5 times as high as the number in the most-fit fourth; men in the two intermediate fourths had corresponding CVD death rates. The risk of dying from a heart attack was about three times greater for the least-fit fourth than for the fittest fourth even after adjusting for smoking, blood pressure and cholesterol levels, so that the link between fitness and health cannot be explained away by genetic differences affecting these characteristics.

The authors conclude that "most of the protective effects of a higher level of physical fitness must act by other mechanisms, at least in part, in addition to [their] well-known effects on blood pressure and serum lipids." They suggest two possibilities: physical training might reduce the risk of heart attack by enhancing blood supply to the heart muscle or by reducing the ease with which blood clots. Results from an unpublished study on women are similar, Ekelund says.

How hard does one have to exercise to gain the presumed benefit? Almost half of the most-fit fourth of the sample reported that they exercised regularly and vigorously (three or more times per week was typical); only 19 percent of the least-fit fourth did so. Ekelund says exercise is likely to benefit most people, whether they are very unfit or already moderately fit: protection increased steadily over the range of fitness in the sample. Men who had evidence of CVD at the outset and were studied separately also benefited from exercise.

The value of exercise is supported by other recent studies. For example, studies of survivors of heart attacks suggest that taking part in an exercise program reduces the risk of a second heart attack. Arthur S. Leon of the University of Minnesota has shown that even a small amount of voluntary exercise can reduce CVD deaths in unfit people who are at high risk.

Ekelund and his coauthor William L. Haskell argue that the American Heart Association and the National Cholesterol Education Program should now classify inactivity as a primary risk factor for CVD. The association lists inactivity as a secondary risk factor and the education program calls exercise an "important adjunct" to a good diet. DeWitt S. Goodman, chairman of the program's expert panel, says that armed with Ekelund's results "we would have made it a stronger comment" but maintains that the evidence is not yet firm enough to list unfitness as a primary risk factor. —*T.M.B.*

Labeled Therapy

A gene-transfer experiment may improve a cancer treatment

An experiment in which genetically engineered cells will for the first time be introduced into human patients passed its major regulatory hurdle in December and should be under way early this year. The experiment was proposed by Steven A. Rosenberg and R. Michael Blaese of the National Cancer Institute and W. French Anderson of the National Heart, Lung, and Blood Institute. It would exploit gene-transfer methods, developed by Anderson, to improve an experimental therapy for intractable cancers that has been developed in the past few years by Rosenberg and his colleagues. At the same time a test of the gene-transfer technique could hasten the day when patients suffering from a grave inherited disease could receive copies of a functioning gene to replace one that is defective or missing.

The experimental cancer therapy is a modification of an earlier approach in which Rosenberg showed that tumors sometimes regress if patients are infused with large numbers of their own immune-system cells that have been grown in the laboratory and treated with interleukin-2, a substance that activates them. Rosenberg is now working with cells called tumor-infiltrating lymphocytes (TIL's), which are extracted from a patient's tumors. The cells are grown in culture for several weeks before being reinfused into the patient along with interleukin-2 and the anticancer drug cyclophosphamide. Rosenberg's latest published results show that nine out of 15 patients who had advanced metastatic melanoma experienced substantial regression of their tumors under this treatment; one of the patients achieved total remission.

Rosenberg thinks the therapy could be made more effective if he knew exactly which types of TIL are best at reestablishing themselves and fighting tumors when large numbers of them are introduced into a patient. It might then be possible to use the effective cells in the therapy rather than a mixture of TIL's. The proposed gene-transfer experiment is intended to identify the effective cells by "marking" some TIL's, before they are infused into the patient, with a gene conferring resistance to an antibiotic. Then, after the cells are infused, various cell samples would be taken and cultured in the presence of the antibiotic. Only TIL's that are progeny of the marked cells (and so would carry the resistance gene) would survive; examining them should show which types of TIL had been most effective.

The experiment carries some risk. Regulatory committees were apprehensive that the hybrid retrovirus serving as a vector to transfer the marker gene into TIL cells, although engineered to be incapable of reproduction, might through recombination reacquire that capability and perhaps some characteristics of its parent viruses, one of which causes leukemia in mice.

Last summer a subcommittee of the National Institutes of Health's Recombinant DNA Advisory Committee insisted on withholding its approval for the proposed experiment until Anderson presented additional details. Later the full committee nonetheless approved the experiment on the basis of an oral presentation (stipulating that the number of subjects should be no more than 10 and their expected life span less than 90 days). NIH director James B. Wyngaarden, sensitive to the controversy, declined to authorize the experiment until both committees had reviewed all data in written form; the subcommittee gave its approval in December. —*T.M.B.*

OVERVIEW

Fusion's Future

Will fusion-energy reactors be "too complex and costly"?

These should be halcyon days for fusion-energy research. Never before have the ill effects of fossil-fuel burning, which spews climate-altering gases into the atmosphere, and of nuclear fission, which creates deadly and long-lasting radioactive waste, been so apparent. Never before have fusion researchers been so sure they can design reactors that generate more energy than they consume. Yet even the most optimistic investigators acknowledge that the forecast for fusion energy is cloudy. "There is no doubt that it can work," says Wolfgang Stodiek of the Princeton Plasma Physics Laboratory, who has been in the field for more than 30 years. "The question is whether it will be too complex and costly."

Scientists once envisioned safe "backyard" fusion reactors burning cheap, abundant fuel. It now appears that generators would require exotic, expensive fuels and would produce significant amounts of radioactivity. Moreover, the smallest fusion generators possible would be comparable in size and complexity to the largest of today's fission reactors—and perhaps much larger. In the near term the U.S. and other governments may be reluctant to spend the billions of dollars needed to build huge prototype generators; in the long term, even if the prototypes work, utilities may choose to invest in smaller-scale, less risky technologies.

Most of the major industrial powers, including the U.S., the Soviet Union, Japan and a consortium of European nations, are pursuing fusion research, but budgets are shrinking. Funding in the U.S., which has generally had the most aggressive program, has been cut nearly in half since its peak in the late 1970's, during the Arab oil embargo, to just over $500 million this year. Inevitably, perhaps, competing investigators have begun to squabble over how their dwindling resources should be spent. "There is a lot of backbiting in this field," acknowledges John F. Clarke, who oversees fusion-energy research for the U.S. Department of Energy, "but people are doing the best they can."

The workers all share the same goal: harnessing the energy that makes stars shine and thermonuclear bombs explode. Unlike fission, which takes place when the nuclei of heavy atoms such as uranium split, fusion occurs when two nuclei—usually of light atoms such as hydrogen or helium—merge into one, releasing energy in the process. Many schemes for inducing fusion have been given consideration over the past 40 years, but attrition and a lack of funding have narrowed the field to two serious contenders: the traditional method, magnetic confinement, and the flashy upstart, inertial confinement.

Magnetic confinement was first attempted in the early 1950's and is still the most widely pursued approach, accounting for more than two-thirds of the U.S. effort and an even greater percentage worldwide. It calls for heating and compressing a plasma (a gas whose molecules have been stripped of their electrons) confined by a magnetic field. The invention by Soviet physicists of a doughnut-shaped vessel called the tokamak marked a breakthrough in the research in the late 1960's. Several tokamaks, including Princeton's Tokamak Fusion Test Reactor and the Joint European Torus in Culham, England, have created plasmas hotter than the interior of the sun, but none has been able to contain a plasma for more than an instant before it squirts free, dropping in temperature and density and becoming contaminated by contact with the containment vessel. Moreover, the machines are at least an order of magnitude away from achieving ignition: the point at which the fusion reaction produces enough heat

to sustain itself without further heating by the machine.

Up until now plasmas have consisted primarily of deuterium, a heavy isotope of hydrogen that is abundant in seawater and is not radioactive. Investigators once hoped deuterium alone would suffice as a fuel, but it has become clear that the fusion of deuterium yields too little energy. Most workers now favor combining deuterium with tritium, the isotope of hydrogen that helps to boost the yield of thermonuclear bombs. Deuterium-tritium fusion can yield up to 200 times more energy than deuterium-deuterium fusion. Although tritium is rare in nature, it can be manufactured in so-called breeder reactors by exposing the element lithium to highly energetic ("fast") neutrons. Since the deuterium-tritium reaction spawns fast neutrons in abundance, a fusion reactor blanketed with lithium could breed its own tritium. Heat from the lithium blanket could also generate steam for driving turbines.

There are drawbacks to this scheme, however. Fuel breeding is not a simple technology (as the current, well-publicized troubles of the U.S. military's breeder-reactors demonstrate); a lithium-blanketed reactor would be particularly difficult to design because lithium is highly reactive, even explosive. Moreover, tritium itself is radioactive, and fast neutrons from the deuterium-tritium reaction would eventually make the entire containment vessel radioactive as well. Although the radioactivity would not be nearly as long-lived or lethal as that produced by today's fission reactors, coping with it—both during and following the reactor's active years—would add considerably to the total operating cost.

Some investigators, notably Bruno Coppi of the Massachusetts Institute of Technology, advocate combining deuterium with helium 3, a nonradioactive isotope of helium. The reaction has a high yield and is clean: it produces no neutrons. Helium 3 is rare on the earth, but like tritium it can be bred by exposing light elements to fast neutrons. Some workers have also proposed mining helium 3 on the moon, where the solar wind apparently deposits large amounts of the substance. The problem is that deuterium-helium-3 ignition would require higher temperatures, and therefore a more powerful and expensive containment vessel, than deuterium-tritium ignition.

The Energy Department plans to build a reactor at Princeton that, it is hoped, would achieve ignition in a tritium-deuterium plasma. Called the Compact Ignition Tokamak, it would cost at least $400 million and might be ready for experiments by 1996. The machine would not be a true prototype: it would achieve ignition for only five or 10 seconds at a time at most, and it would have no equipment for converting the fusion energy into electricity. Meanwhile the U.S., the U.S.S.R., Japan and Europe are considering banding together to build a full-scale prototype generator, called the International Thermonuclear Experimental Reactor (ITER), that could be operating by early in the next century. An international team is supposed to present a design by the end of 1990, when the participants will decide whether to proceed.

Already there is disagreement over the ITER's basic design. Most members of the design team favor building a high-performance reactor that has superconducting magnets and would generate some one billion watts of thermal energy, as much as a typical large fission reactor. A machine with a greater output or size, workers say, would appear impractically large to most governments and utilities. But Paul H. Rebut, the chief designer and now the director of the Joint European Torus, considered by some to be the most advanced tokamak in the world, advocates building a very big, three-billion-watt reactor incorporating conventional, nonsuperconducting magnets. A smaller, less conservative reactor, Rebut says, might not work. For the U.S. the issue might turn out to be moot. According to some observers, the U.S. is not likely to join in actually building the international reactor, both because of its projected cost (at least $2 billion) and because of concerns about the transfer of advanced U.S. technology to other countries, particularly the Soviet Union.

The field of inertial confinement, a spinoff of laser research in the mid-1960's, is even more wracked with technical and political disputes. Whereas magnetic confinement seeks to mimic the sustained conflagration of a star's interior, inertial confinement aims to re-create a thermonuclear explosion, in miniature. The idea is to blast a small, spherical capsule of deuterium and tritium from all sides with radiation; the resulting implosion makes the fuel so dense and hot that it ignites. The official goal of the U.S. inertial-confinement program is not to produce a new source of energy but to help workers study weapons-related physics; the Government has classified many aspects of the program and has prohibited any significant international cooperation. Most inertial-confinement workers contend nonetheless that their own primary goal is producing what one refers to as "energy for the planet."

In the field's giddy early days some researchers predicted that only a few hundred joules of laser light (a 100-watt bulb generates 100 joules per second) would be needed to achieve ignition. The estimates have escalated over the years, as have doubts about the technology's feasibility. Recently, however, the field has been revitalized by the results of secret experiments, code-named Centurion-Halite, done at the Nevada Test Site. In the experiments, X rays from underground nuclear explosions ignited capsules of tritium and deuterium. Proponents of inertial confinement say the experiments provide something that has eluded magnetic-confinement workers: a proof of principle. All that needs to be done now, they assert, is to build a "driver," or radiation source, that can produce the same results. The Energy Department is considering building a so-called Laboratory Microfusion Facility in the late 1990's that would do just that, according to David N. Bixler, an official in the department's inertial-confinement division.

A number of laboratories are competing for the honor of designing the facility's driver. It is still unclear how powerful the driver should be and what form it should take. According to most investigators, the Centurion-Halite tests indicate that no more than 10 million joules of energy will be necessary to ignite targets. That is more than 100 times as much energy as the world's largest laser—the neodymium-glass Nova at the Lawrence Livermore National Laboratory—can deliver. Erik Storm, who oversees inertial-confinement research at Lawrence Livermore, contends that a scaled-up version of the Nova would make the best driver for the microfusion facility. He notes that the Nova has had much more success than any other aboveground driver at compressing and heating targets.

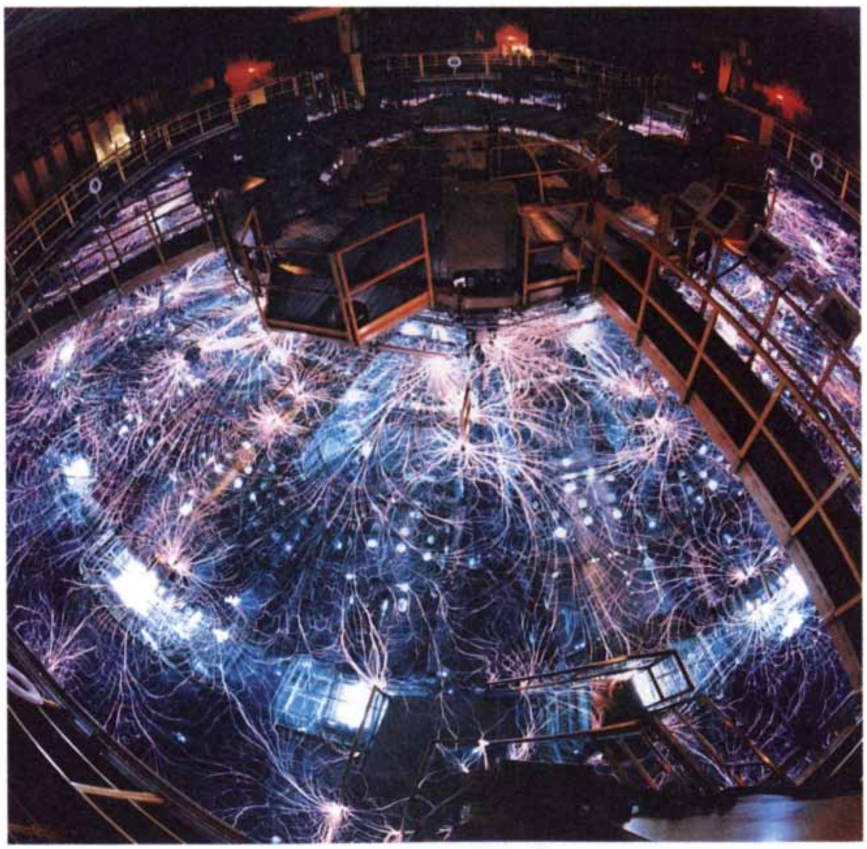
Workers at other laboratories are quick to point out that glass lasers are notoriously inefficient (the Nova, for example, converts less than 1 percent of the electrical energy it consumes into light energy) and subject to overheating. Glass lasers are also expensive: the Nova cost more than $200 million, and a glass laser 100 times as powerful might easily cost 10 times as much, according to some estimates. The Energy Department is considering several other drivers, including a krypton-fluoride gas laser at the Los Alamos National Laboratory, a light-ion accelerator at the Sandia National Laboratories and a heavy-ion accelerator at the Lawrence Berkeley Laboratory, but all are in preliminary stages of testing and have drawbacks of their own, according to Bixler.

An alternative to all these approaches has been proposed by a former physicist at Los Alamos. Taking a position strikingly similar to Rebut's in the debate over the ITER, P. Leonardo Mascheroni says those who maintain that 10-million-joule pulses will ignite targets are interpreting the Centurion-Halite data too optimistically. If the microfusion facility's driver cannot deliver 10 times as much energy, according to Mascheroni, it might never achieve complete ignition. The most economical way to achieve 100-million-joule pulses, he says, is with a hydrogen-fluoride laser. Mascheroni, who was fired by Los Alamos early last year—for not supporting the laboratory's official line, he contends—is now taking his case directly to Congress and other Government officials.

Still another dispute concerns how best to achieve the highly uniform irradiation of the target that is needed to induce ignition. A technique known as direct drive positions numerous beams symmetrically around the target. In a method called indirect drive the target is suspended in the middle of a cavity lined with a foil made of gold or some other heavy metal. A beam of radiation enters an aperture in the cavity and vaporizes the metallic lining, generating X rays that converge on the target. Because a similar technique is employed in triggering fusion in thermonuclear bombs, details about indirect drive are classified, but advocates maintain that the Centurion-Halite tests and other experiments indicate it is the best way to achieve ignition.

Direct-drive enthusiasts, including workers at the University of Rochester's Laboratory for Laser Energetics and at the Naval Research Laboratory in Washington, demur. While acknowledging past difficulties at compressing targets uniformly, they argue that recent advances may soon enable them to equal indirect drive's results. They point out, furthermore, that direct drive, because it does not require the conversion of the primary radiation source into X rays, delivers energy much more efficiently than indirect drive does and therefore would reduce the power requirement of a driver.

Unlike fission reactors, which can ride a satellite, fusion reactors must be gigantic machines



***ELECTRICAL DISCHARGES** lace the surface of water bathing the Sandia National Laboratories' Particle Beam Fusion Accelerator II as it fires a pulse of light ions. The Department of Energy is considering employing such a machine as a "driver" for an advanced inertial-confinement facility. Its million-joule beams are more powerful than any laser's, but they are also much more difficult to focus.*

When—or if—these disputes are resolved, the field still faces some staggering technical problems. Assuming that Mascheroni is wrong and that a 10-million-joule driver would be sufficient to achieve ignition, and assuming that ignition yields 100 times as much energy as the triggering radiation (the stated goal of Lawrence Livermore), then each firing would yield one billion joules of energy—the amount produced by the explosion of 500 pounds of TNT. The Laboratory Microfusion Facility would probably generate only one such event a day, according to Bixler, whereas an actual generator would probably have to fire five or six times per second. What kind of vessel could withstand such violent, repeated shocks? Various schemes have been proposed, including one that calls for surrounding the reactor's core with "waterfalls" of liquefied lithium. But then, as with magnetic confinement, the question of scale arises. A reactor that generated, say, five billion-joule shots per second would dwarf even the behemoth proposed by Rebut.

Energy Department officials insist that until a clearly superior energy source emerges research on both magnetic and inertial confinement should and must continue. Clarke even predicts that, given sufficient Government support, magnetic-confinement reactors could come on-line by 2020 and inertial confinement could provide a competitive alternative a decade or two later. What if someone invents a cheap, efficient solar cell? "I guess that would be it," he replies. *—John Horgan*

SCIENCE/SCOPE®

Hughes Aircraft Company was the recipient of the U.S. Army Award for outstanding achievements in Value Engineering for cost-savings. In 1987, Hughes Value Engineering Change Proposals (VECPs) saved customers over 53 million dollars. Since 1964, 735 Hughes VECPs on 53 programs have resulted in a total savings of 1.136 billion dollars. Some of the programs benefitting from the VECPs include F/A-18 avionics, Maverick, Phoenix, and Advanced Medium Range Air-to-Air Missiles, M1 Abrams Tank, Bradley Fighting Vehicle, and the AN/UYQ-21.

A new Space Based Radar Program will involve the placement of a constellation of sensor platforms in the Earth's orbit between 600 and 6,000 nautical miles in altitude for wide area surveillance of ships, aircraft, and cruise missiles. Hughes, as a member of the Grumman-led team, will define technology requirements and an implementation plan for the radar RF and processing sections, which will interface with Grumman's SBR system. An operational demonstration-validation phase will lead to a first launch in the mid-1990's. The Space Based Radar Program is a joint U.S. Air Force and Navy program.

An infrared sensor system for the joint services' V-22 Osprey aircraft will help crews navigate and land in darkness and during periods of poor visibility. The Infrared Detection Set (IDS), developed by Hughes, senses small differences in radiated heat and provides a TV-like image of the surrounding area. The V-22, the first of the new tilt-rotor aircraft for the U.S. Armed Forces, is able to take off and land like a helicopter, but fly like a plane. Hughes will start flight testing the IDS in early 1989, with flight test support continuing through 1991.

A weather satellite in geosynchronous orbit above the Atlantic seaboard "sees" cloud cover even at night. The Geostationary Operational Environmental Satellite (GOES) H, built by Hughes, sees through a combination telescope and sensing instrument called a visible-infrared spin scan radiometer atmospheric sounder (VAS). Developed by Hughes, the VAS uses reflected visible light and infrared thermal radiation to gather images of Earth and its atmosphere. The last in a series of five weather satellites, GOES H watches for storms that threaten the East coast.

Hughes Missile Systems Group is advancing every phase of research and development, as it applies to tactical guided missile systems and strategic defense. Opportunities for engineers and scientists are in analog/digital circuit design; high-voltage power supply design; electro-optical design; IR imaging sensors; focal plane arrays; and systems engineering. Send resume to Hughes Engineering Employment, Dept. S2, 8433 Fallbrook Ave., Canoga Park, CA 91304. Equal opportunity employer. U.S. citizenship may be required.

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

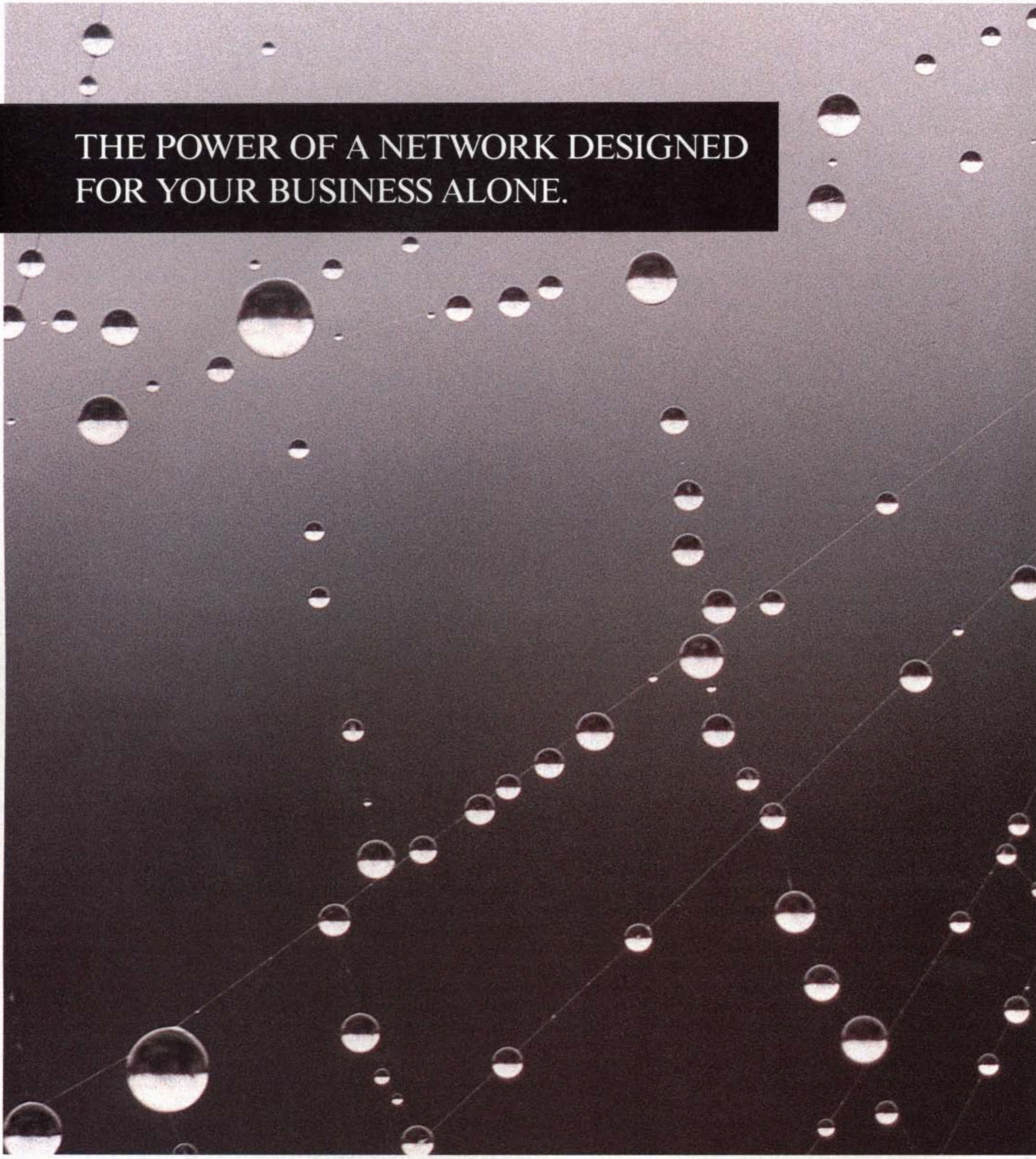
HUGHES

Subsidiary of GM Hughes Electronics

For the growth of your business, the prosperity of your business, there is no more powerful tool than the telecommunications network.

And at GTE, there is nothing vague about the power we are putting into your hands when we design that network for you.

For one of the nation's leading retailers, we are now implementing one of the largest satellite-based private



SCIENTIFIC AMERICAN February 1989 Volume 260 Number 2

Soviets in Space

Cosmonauts have spent more than 5,600 days on board Soviet space stations since 1971. Yet cosmonaut activities are just a small part of the Soviet Union's robust space program

by Peter M. Banks and Sally K. Ride

Last November the Soviet Union launched its space shuttle *Buran* from the Baikonur Cosmodrome near Tyuratam. *Buran* (the Russian word for *snowstorm*) was lifted into orbit by the world's largest rocket, circled the earth twice and then executed an apparently flawless automated reentry and landing, touching down on a runway on the steppes of central Asia. It carried no crew, no satellites and no experiments, but its brief maiden voyage was a resounding technological success.

Buran's flight was the 77th Soviet space launch in 1988. The shuttle was originally scheduled to be launched on October 29, but the countdown was aborted 51 seconds before lift-off when an access arm on the launch pad failed to retract. During the next three weeks, while engineers readied *Buran* for a second launch attempt, Soviet space activities continued apace. The Soviet space station *Mir* circled the globe 330 times, the two cosmonauts on board the station broke the space endurance record of 326 days and the robotic Soviet spacecraft *Phobos* sailed ever closer to Mars.

Buran and its huge launch vehicle Energia are the latest additions to a wide-ranging space program that has successfully lofted eight space stations since 1971 and has maintained a regular if not continuous human presence in space for the past 12 years. Although they are impressive, these feats represent just a small part of the Soviet space effort.

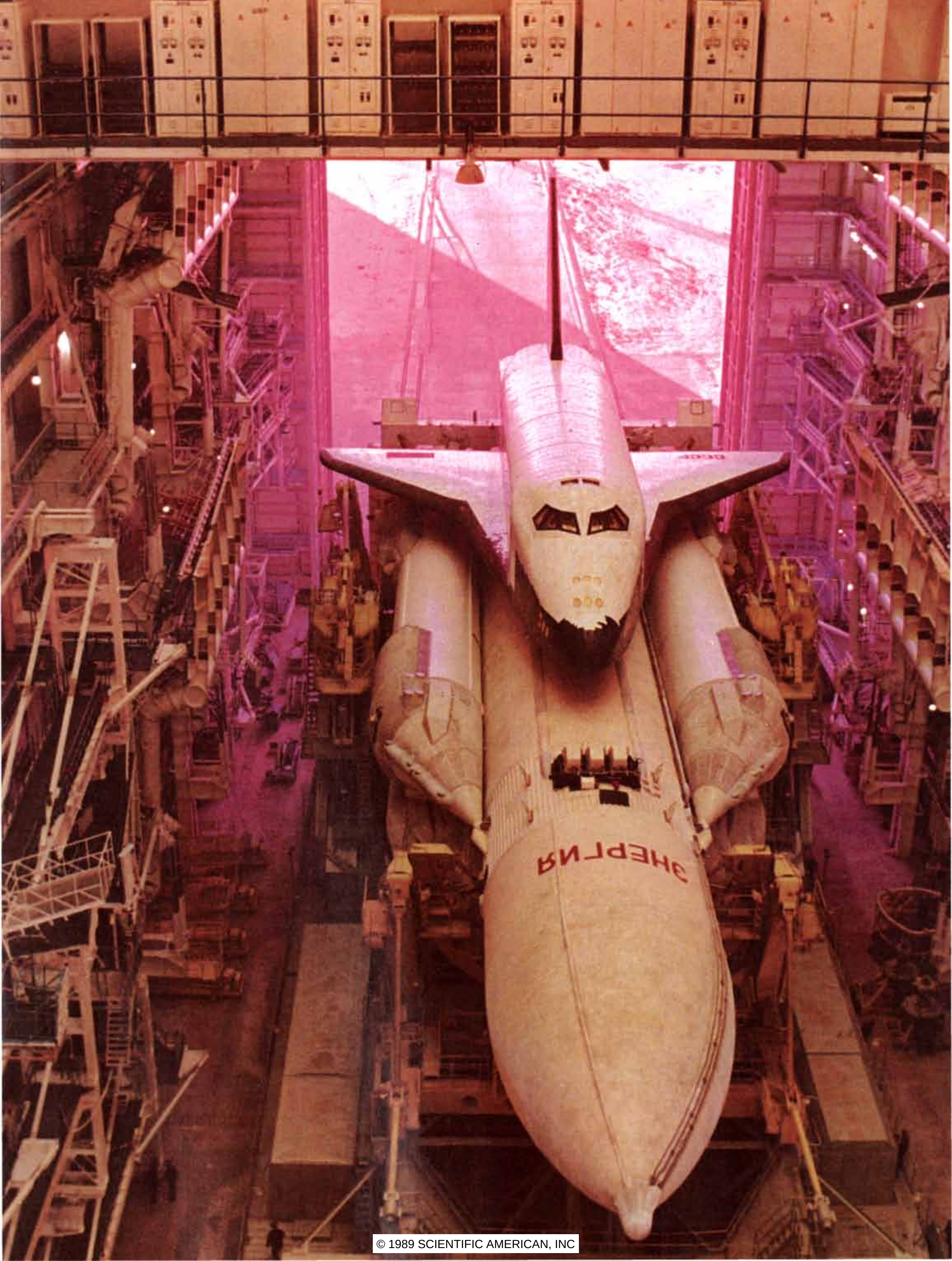
Over the past three decades, with designs that emphasize existing technology and an evolutionary approach, the Soviets have built the infrastructure and capability to achieve and sustain a vigorous human presence in space. The magnitude and scope of this program is only now being appreciated in the West.

Before 1985 much of what was known about the Soviet space program was the result of painstaking detective work by a small group of dedicated observers in Europe and the U.S. Little information was available from official or firsthand sources. The Soviets remained reticent about their activities as a matter of policy, and the U.S. Government elected to withhold the information acquired through its own channels. The result was a general state of ignorance about important space-related events and developments taking place within the U.S.S.R.

That situation, at least as it concerns civil space activities, has changed dramatically. In the past three years the Soviets have come forward with many details about their space ventures. They have described with considerable candor the circumstances of cosmonaut Yuri A. Gagarin's death in an airplane crash in 1968, and they have released important technical details of their Proton launch vehicle, the first rocket designed exclusively for use by the space program. The Yuri Gagarin cosmonaut training center at Star City (near Moscow) has been opened to Western visitors, and several space launches, including that of *Buran,* have been announced in advance; some have even been shown live on Soviet television. Last fall, when two cosmonauts encountered dramatic problems in attempting to return to the earth, the Soviets resolved the situation in full view of the world.

PETER M. BANKS and SALLY K. RIDE began collaborating on a report of Soviet piloted space missions soon after Ride moved to Stanford University in 1987. Banks is professor of electrical engineering at Stanford and the director of the university's Space, Telecommunications and Radioscience Laboratory. He became interested in the U.S.S.R.'s space-station program through his experience leading a National Aeronautics and Space Administration advisory committee that helped to define the scientific uses of the U.S. space station. Ride is a science fellow at Stanford's Center for International Security and Arms Control. She flew on two space-shuttle flights in 1983 and 1984 and was in training for her third at the time the space shuttle *Challenger* exploded. Ride terminated training to serve on the presidential commission investigating the accident, then accepted a position as special assistant to NASA's administrator for strategic planning before going to Stanford.

SOVIET SHUTTLE *BURAN,* mounted on the Energia booster that launches it, rests in a hangar before making its debut flight this past November. The design is obviously similar to that of the U.S. space shuttle, but *Buran* does not have to be piloted. The Soviets say they have at least one other shuttle in production.


ЭНЕРГИЯ

The picture that has since emerged in the spirit of glasnost is one of a well-supported national enterprise the sheer magnitude of which is overwhelming. In 1987 launches from Soviet facilities accounted for 86 percent of space launches worldwide. Whereas the U.S. launched eight expendable rockets in that year, the Soviets launched 95. And 1987 was not an exceptional year: the Soviets have been launching almost two rockets per week since the mid-1970's.

In spite of the disparity in launch rates, the U.S. and the U.S.S.R. have about the same number of active satellites in orbit, and these satellites are engaged in similar activities. It is estimated that about 80 percent of all Soviet space launches are related in some way to programs for national security: space-based reconnaissance, communications, navigation and missile early-warning satellites. About 10 percent are related to the support and operation of the space station. The remaining launches are devoted to automated scientific satellites, civil-communications satellites and navigation satellites.

One particular strength of the Soviet program is its diverse collection of trustworthy rockets. The Soviets currently have nine different launch vehicles in addition to the new Energia booster. Six of these are direct descendants of Soviet ballistic missiles and have been in use for decades. All except Energia use well-understood liquid-fuel rocket technology; the new rocket uses advanced liquid hydrogen-oxygen propellants. The Soviets rarely retire a launch vehicle that has been proved, and so they have accumulated considerable experience with all but the youngest of their boosters.

The SL-4 has been the workhorse of the program for nearly 25 years. More than 900 SL-4's have been launched since 1964, including 43 in 1987, and it is still the only booster the Soviets use to lift their cosmonauts into orbit. The Proton rockets, SL-12 and SL-13, were first tested in 1965; more than 115 of them have been launched since then. The Proton can lift more payload to low earth orbit than any current U.S. expendable booster. Both the SL-4 and the Proton are integral parts of the space-station project: the SL-4 launches the Soyuz capsules that carry cosmonauts to the station, and the Proton launches the modules from which the space station is built.

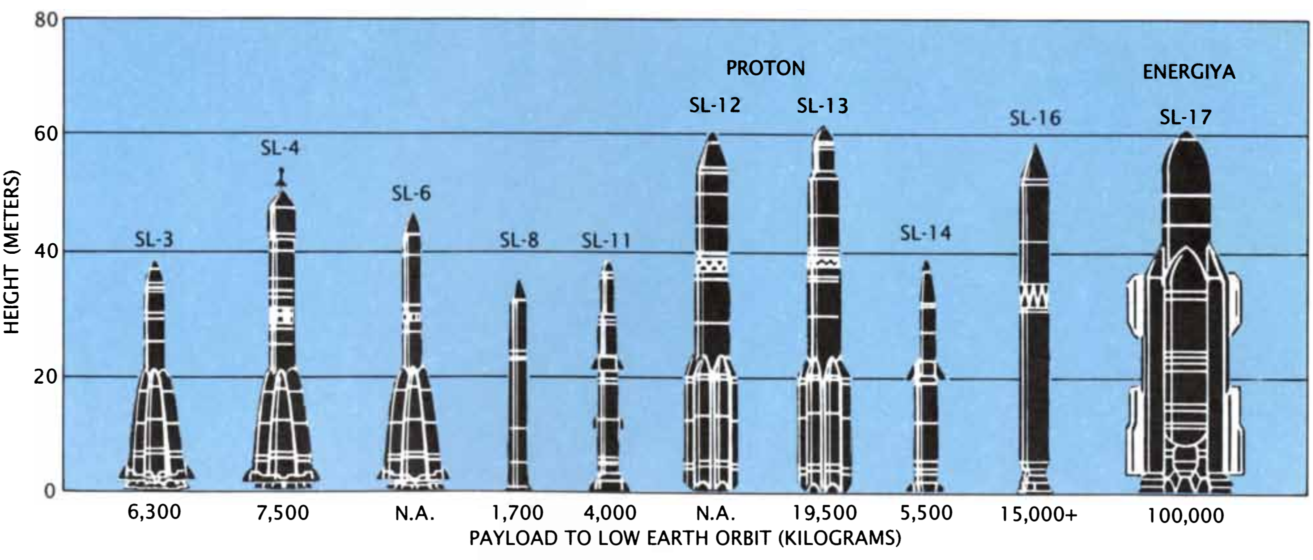
Although the rocket systems developed in the 1960's are still the backbone of Soviet space transportation, the development of launch vehicles is not stagnant. The SL-16 was introduced in 1985, and Energia was tested for the first time in 1987. Energia can launch payloads of about 100,000 kilograms into low earth orbit. Its design allows for significant flexibility: it can transport into space either the Soviet space shuttle or a large cargo canister.

Energia is closely related to the SL-16 rocket; indeed, four SL-16 first-stage boosters are used on Energia as strap-on first-stage rockets. That design strategy exemplifies Soviet engineering practices, in which more powerful and sophisticated systems typically evolve from modest, proved designs.

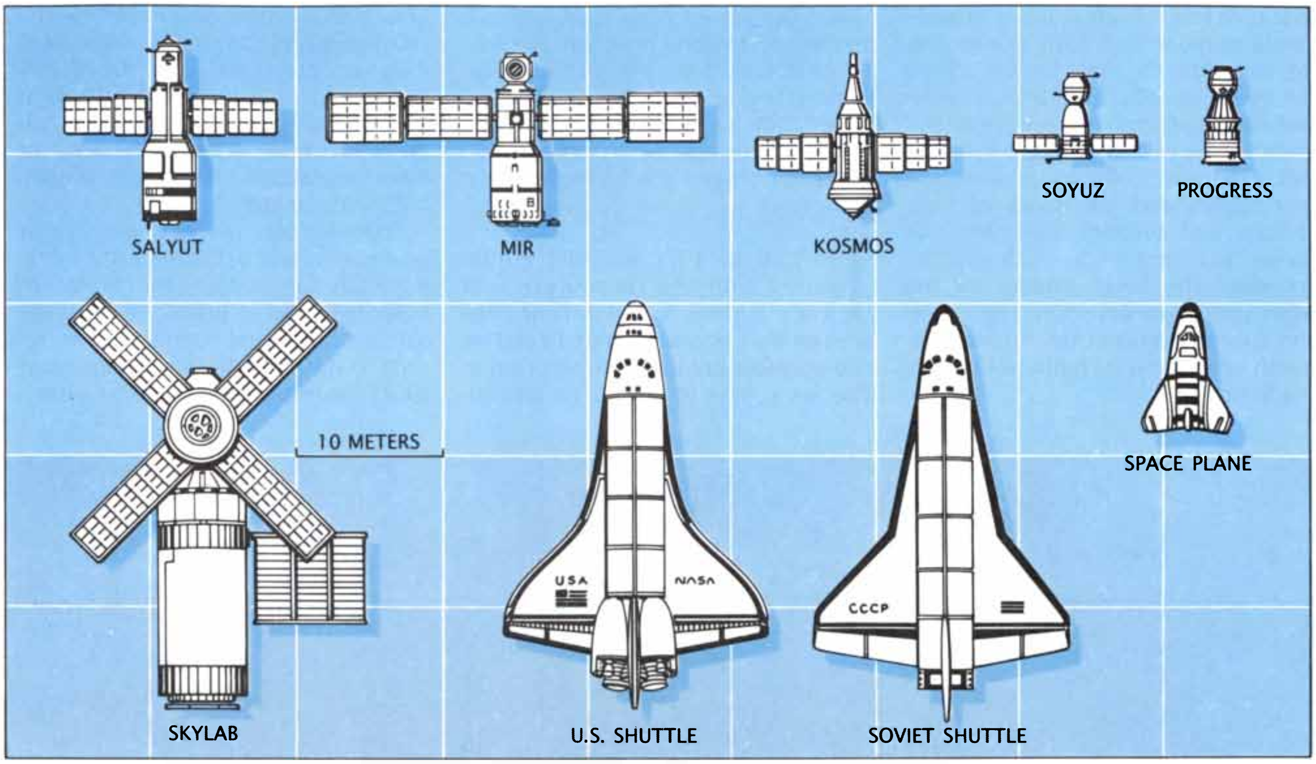
There are two major spaceports in the Soviet Union: Baikonur in central Asia and the Plesetsk missile test range about 800 kilometers north of Moscow, near the White Sea. In 1987 alone the Soviets conducted 47 launches from Baikonur and 47 from Plesetsk. Cosmonaut flights and other flights associated with the space-station program are staged exclusively from the Baikonur facilities, but in 1987 these accounted for only about a fourth of the traffic through that busy spaceport.

An intricate network of facilities throughout the Soviet Union supports space activities by means of satellite and rocket design and production, payload integration and crew-training, launch and mission operations. The network requires experienced individuals and seasoned organizations. Although the pace of technological development in the Soviet Union is slow by U.S. standards, 30 years of consistent commitment and nurturing have produced a robust space-transportation system that can easily accommodate the maintenance of the *Mir* space station as just one of its many other activities.

The age of space exploration began in 1957 with the launch of *Sputnik I*. Although Soviet and American experts had long appreciated the strategic advantages of rockets and earth-orbiting satellites, it was Sputnik that shocked the world into



ARRAY OF LAUNCH VEHICLES used in the Soviet space program evolved from the SL-3 to the Soviets' latest and most powerful addition, the SL-17 Energia rocket. The SL-4 and Proton boosters are the backbone of the space-station program; more than 900 SL-4's and more than 100 Proton rockets have been launched since the vehicles were designed in the early 1960's.



SPACECRAFT that have played or will play historic roles in the Soviet space-station program are shown alongside the U.S. space station *Skylab* and the U.S. shuttle. The Soviets have not acknowledged the existence of the space plane (*lower right*).

an awareness of the significance of access to space. With the launch of Sputnik, the Soviet Union was suddenly perceived as a technological superpower. Although it was far from the truth, this perception fueled American fears, colored political decisions and initiated the space race.

The desire for national security shaped the early space programs of both nations. So did the desire for national prestige. Flights of cosmonauts and astronauts were widely publicized and soon became a measure of each nation's technical prowess. In the early 1960's the Soviets recorded a string of firsts: the first man to orbit the earth (Gagarin), the first woman in space (Valentina V. Tereshkova), the first spacewalk.

But President John F. Kennedy galvanized the U.S. space program and captured the world's imagination by declaring a new long-range goal: to land Americans on the moon before the end of the decade. Kennedy's patriotic challenge and Lyndon B. Johnson's leadership rallied scientific and engineering expertise, gave purpose to the Mercury program and led to the creation of both the Gemini and the Apollo programs.

A journey to the moon was impossible without substantial new capabilities and extensive preparation. While astronauts developed rendezvous and docking techniques and flight surgeons studied the physiological effects of weightlessness on the human body, engineers designed and tested the Saturn heavy-lift launch system and scientists sent space probes a quarter of a million miles to characterize the lunar surface.

At the same time, although they never officially acknowledged it, the Soviets embarked on a parallel effort aimed at putting cosmonauts on the moon. Soviet interest in the project had been evident as early as 1959, when three Luna space vehicles were sent toward the moon, thereby providing, among other things, the first pictures of the moon's far side. Luna spacecraft continued to probe the moon and its environment in 21 more missions.

By the mid-1960's the Soviets had developed their equivalent of the Apollo command module, the Soyuz space capsule. With this versatile vehicle the Soviets learned crucial rendezvous and docking skills. But the Soviets' plans were critically delayed by their inability to develop a heavy-lift launch vehicle like the American Saturn-5 rocket, which could bring almost 140,000 kilograms into low earth orbit. The Soviet hope of being the first to put human beings on the moon evaporated when *Apollo 11* astronauts landed there in 1969.

Although the Soviets engaged in robotic exploration of the moon until 1976, they abandoned their efforts to send cosmonauts there. But cosmonaut activities in low earth orbit continued. Five Soyuz missions were flown in 1969, and in 1970 two cosmonauts managed to recapture the duration-in-space record from the U.S. Gemini program by staying in orbit for 18 days.

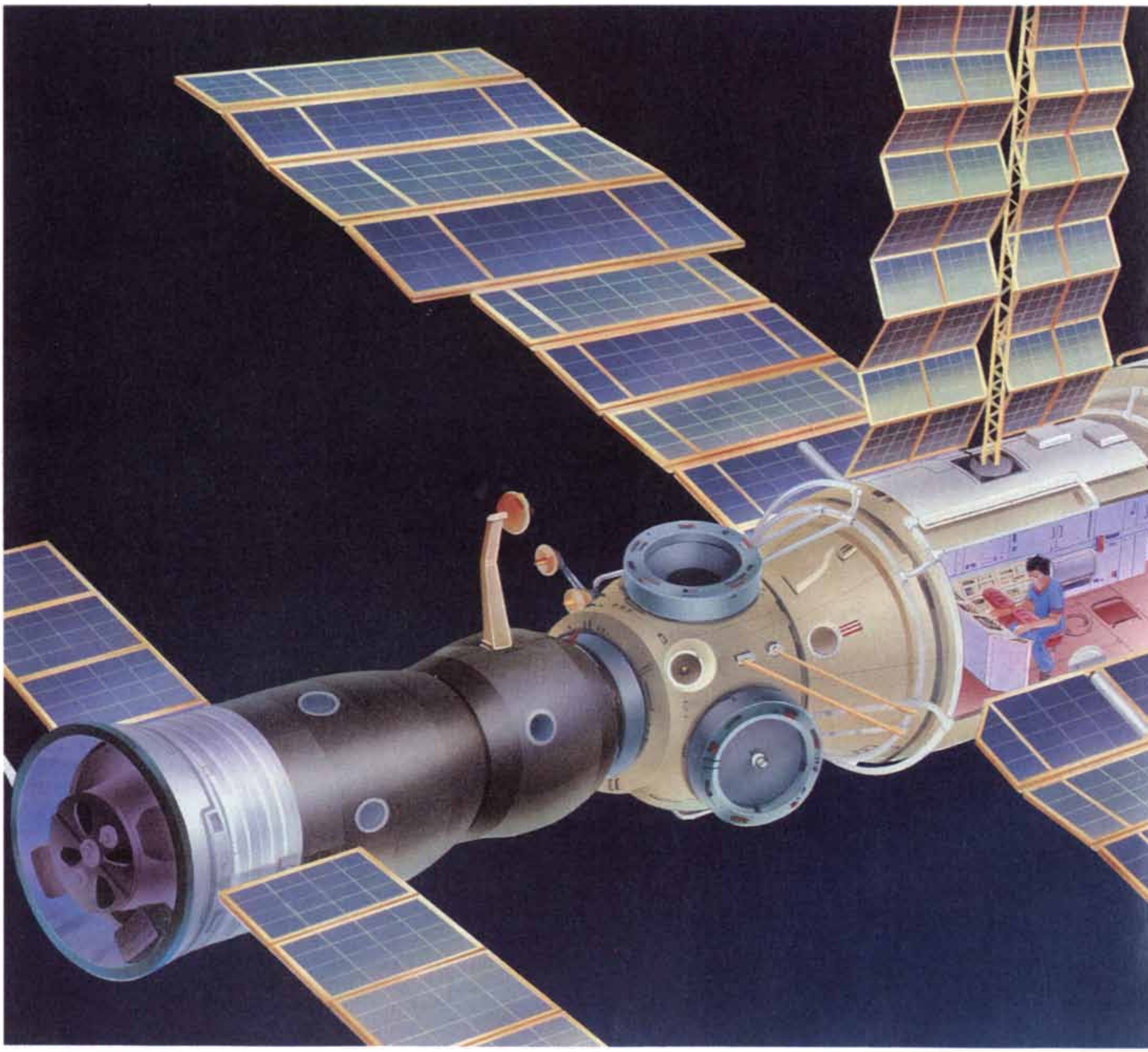
Meanwhile both the U.S. and the Soviet Union continued to develop the technology and infrastructure to support their national security needs in space. Differing technological capabilities led the two countries down different paths. U.S. engineers built progressively more reliable and capable satellites that did not need to be replaced as often as older satellites, and as a result the U.S. launch rate began to decrease in the late 1960's. The infrastructure of the U.S. space program became geared to launching a relatively small number of increasingly sophisticated satellites.

Soviet satellite technology did not progress as rapidly. Consequently the Soviets found it necessary to continue to launch less reliable and less capable satellites more frequently. It was during this time that the Soviets sowed the seeds for what would become an enduring, dynamic space-station program. In the early 1970's the U.S. was riding the crest of the successful Apollo program, and the needs of both civilian and military programs for rocket launches were gradually decreasing. The Soviet Union, on the other hand, was developing an extensive space-support infrastructure, primarily to support its national-security requirements.

And the Soviets continued to develop their cosmonaut program. The Soyuz capsule and the Proton rocket developed for the lunar program became instead the building blocks of a program that sought to sustain human beings safely and productively in low earth orbit.

In April of 1971 a Soviet Proton rocket launched the world's first space station, *Salyut 1*. Four days later an SL-4 boosted *Soyuz 10* and its two cosmonauts into low earth orbit. The Soyuz was to deliver its passengers and a small cargo load to the station and remain docked there until it was needed to transport the cosmonauts back to the earth. Although primitive by U.S. standards, the Salyut constituted major progress toward establishing long-term human habitation in space.

Unfortunately the first attempt at occupation was a failure: after docking with the station, the *Soyuz 10* cosmonauts were unable to open the Salyut's hatch and were forced to return to the earth. Six weeks later *Soyuz 11* made a flight that, although ini-



***MIR* COMPLEX shows a typical configuration of the versatile space station. The Soyuz capsule that delivers the station's crew is anchored at *Mir*'s aft multidocking port; that port still has four docking sites free (*left*). The *Kvant* science module is**

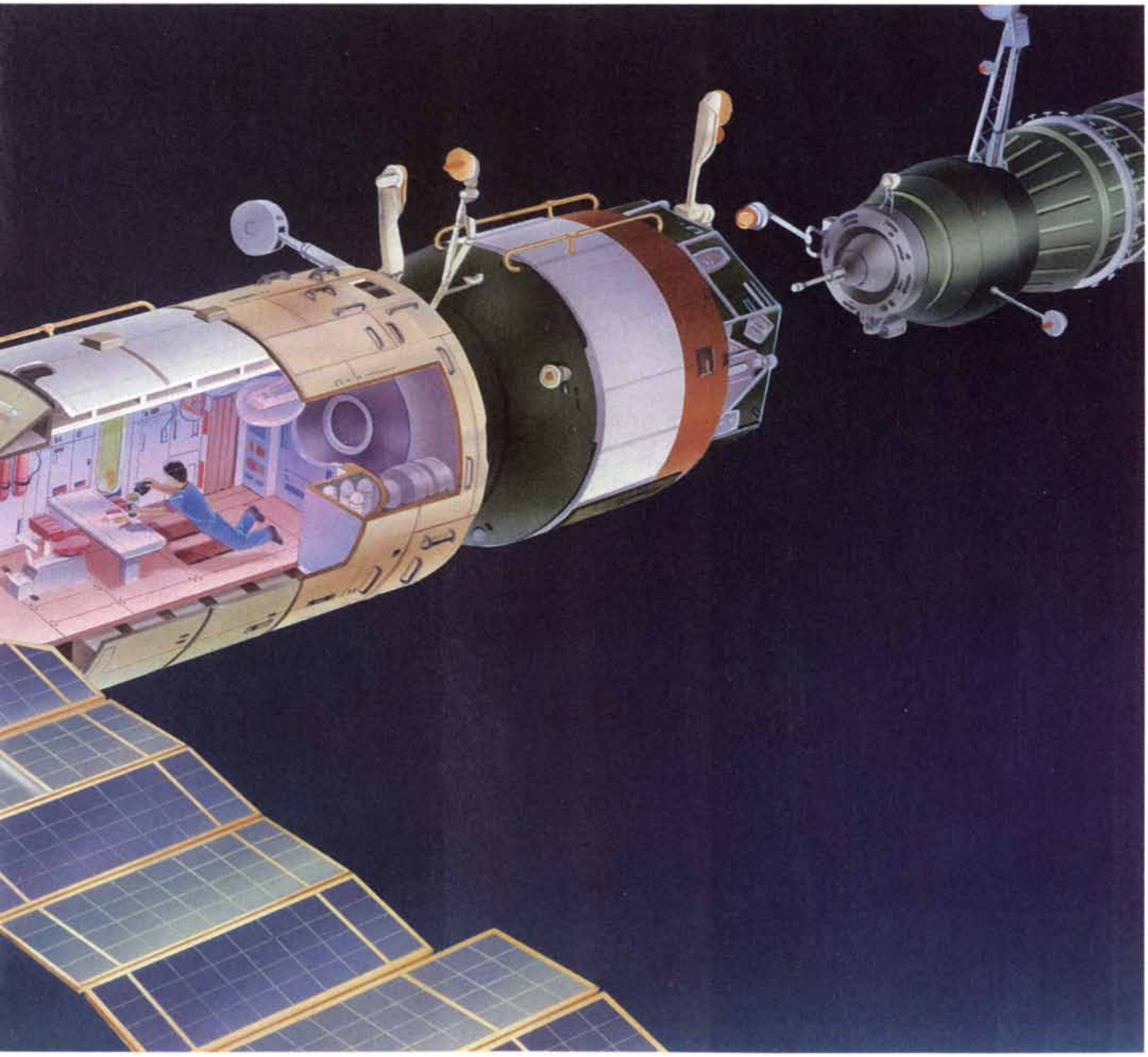
tially more successful, ended in disaster. Three cosmonauts docked and boarded *Salyut 1* and stayed for a record 23 days in space. Then during the return the Soyuz capsule lost its pressure, and the crew died before touchdown.

The accident prompted a 27-month moratorium on Soyuz flights and an extensive redesign of the Soyuz spacecraft. The Soviets also decided that henceforth cosmonauts would wear pressure suits during launch and reentry, a decision that still stands.

Salyut 1 was deliberately removed from orbit during the Soyuz stand-down. Two attempts to replace it two years later were unsuccessful. *Salyut 2*, which was launched in April of 1973, could not be stabilized and eventually broke apart in orbit. Another spacecraft launched in May of the same year apparently had attitude-control problems and expended all its propellants.

Three Salyuts launched between June, 1974, and June, 1976, were more successful; although there were still significant problems, each station housed at least one cosmonaut crew. The stations were occupied only from 10 to 15 percent of the time they were in orbit. On board the cosmonauts carried out a variety of basic experiments and observations, many of which Western analysts believe were sponsored by the military.

The four successful Salyut stations launched between 1971 and 1976 can be characterized as the first generation of Soviet space stations. Their size was modest: they had about 100 cubic meters of habitable volume and weighed 26,000 kilograms with a Soy-



attached to *Mir*'s fore docking port and an automated Progress cargo ferry is preparing to dock at the fore of *Kvant* (*right*). Inside *Mir* there are accommodations for two resident cosmonauts and for visiting crews having up to four people.

uz capsule attached. Electrical power was in short supply and the accommodations were spartan. In addition, first-generation Salyuts had only one docking port, and since the Soyuz occupied that port while cosmonauts were on board, it was impossible to resupply the station. Each Soyuz had to bring along enough food, water and propellants to sustain its two- or three-member crew during its visit.

Even with those limitations, the Soviets gained considerable experience in operating habitable facilities in low earth orbit. Meanwhile the U.S. had also been experimenting with long-term habitation of low earth orbit. The *Skylab* space station was launched in 1973 by the Saturn-5 rocket that had been developed during the Apollo program. *Skylab* was far more commodious than the Salyut space stations: it had more than 350 cubic meters of living space and weighed 90,000 kilograms including the Apollo module that delivered the crew.

Three separate astronaut crews of three members each visited *Skylab* for periods of 28, 59 and 84 days. They undertook a variety of experiments in space medicine, human physiology and materials physics and made solar and astronomical observations. Although *Skylab* had two docking ports, no attempts to resupply it were ever made. In 1974 it was put into orbital hibernation, and five years later atmospheric drag brought the space station to a fiery end in the atmosphere over Australia.

In September, 1977, the second generation of Soviet space stations was introduced with the launch of *Salyut 6*. The central core of the new station was virtually identical with those of earlier stations, but it had several improvements, the most important of which was the addition of a second axial docking port. This feature enabled two vehicles to visit *Salyut 6* simultaneously, and so supplies and fuel could be brought up to the station while a crew was already on board.

The Soviets sent the first supply ferry to dock with *Salyut 6* in 1978. The ferry was an automated Progress vehicle, a modified Soyuz that is also launched by an SL-4 and is guided to its rendezvous with the Salyut by flight controllers on the ground. Once it is attached, a Progress tanker can transfer fuel and liquids without the assistance of the cosmonauts. Progress spacecraft are still the primary means of resupplying Soviet space stations.

The automated supply flights of Progress uncoupled cosmonaut and equipment transport to Salyut and thus cleared the way for longer cosmonaut visits. The first *Salyut 6* crew stayed in orbit for 96 days, shattering both the Soviet and the American record of 63 and 84 days respectively. In 1982 *Salyut 7* replaced *Salyut 6*, and the space-endurance record was systematically extended. Successive resident crews stayed in orbit for 96, 140, 175, 185, 75, 211, 150 and finally 237 days. Taken together, *Salyut 6* and *Salyut 7* were staffed nearly 45 percent of their time in orbit and housed 11 long-duration crews and 15 visiting crews, including 10 foreign cosmonauts.

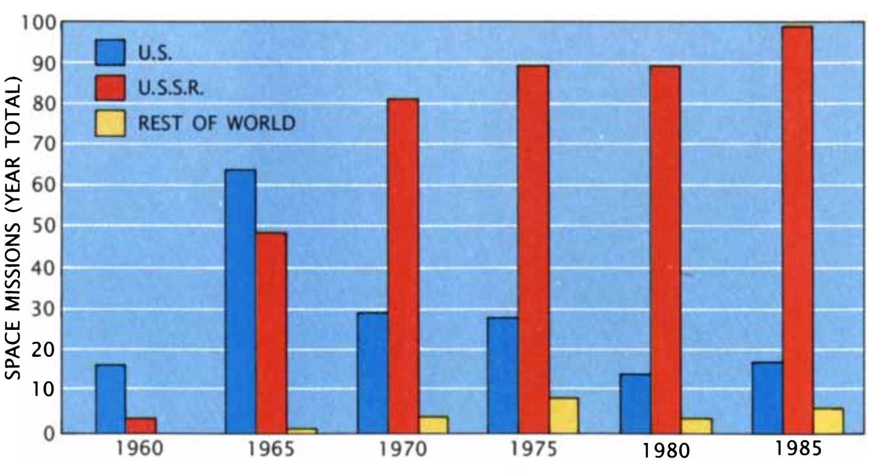
During *Salyut 7*'s tenure Soyuz modules and Progress vehicles made 52 dockings. Automatic rendezvous, docking and refueling in orbit became an established and routine capability. The Soviets introduced a new Soyuz module, the Soyuz-T, and there were four flights of a new multipurpose vehicle that could be used to carry cargo, expand space stations and serve as a space tug. *Salyut 7* was equipped with a redesigned docking adapter and additional solar panels to provide more electrical energy. It is clear the Soviets were learning from experience.

The *Mir* space station, which was launched in February of 1986, represents the third generation of Soviet space stations. *Mir*'s genealogy is readily apparent—it looks much like the Salyuts—but significant improvements inside and out make *Mir* qualitatively different from its predecessors. Eight separate computers provide a more modern operational environment. Automatic systems now control most life- and orbital-support systems, freeing cosmonauts from time-consuming "housekeeping" tasks. Larger and more efficient photoelectric panels supply the station with more electrical power. Communications to the cosmonaut flight-control center near Moscow can be routed through a Soviet communications relay satellite, or in the traditional manner, through tracking ships and ground stations located within the Soviet Union.

The most significant advance, however, is the addition of a five-port docking adapter at one end of the space station. This simple fixture adds a powerful flexibility to the basic Salyut design and greatly enhances the capabilities of the Soviet space-station program. Along with the port at the station's fore, the aft axial port brings up to six the number of spacecraft that can be attached to the station at one time. The latest Soyuz capsule (the Soyuz-TM), the older Progress tankers and cargo ships and the multipurpose Kosmos vehicles can all mate with *Mir*.

The Soviets began to exploit the growth potential of *Mir* in April of 1987, when they added the 11,000-kilogram science module *Kvant*. A Proton rocket launched *Kvant* into low earth orbit. The docking, however, did not go entirely as planned: on the first rendezvous attempt *Kvant* sailed past *Mir* just 10 meters from the station.

On the second try the two spacecraft



NUMBER OF SPACE MISSIONS sponsored annually by the U.S. and the U.S.S.R. has undergone a dramatic reversal since the 1960's. Today Soviet launches account for more than 85 percent of launches worldwide. Because Soviet satellites are short-lived, however, the number of satellites the two nations have in orbit is comparable. China, Japan and the European Space Agency also stage several launches each year.

joined in a loose "soft dock," but they could not be brought together in a solid "hard dock." Finally, as a last resort, the *Mir* crew donned spacesuits and went outside in order to examine the vehicles' docking mechanisms. They soon found that a plastic trash bag was obstructing the hard dock. On consulting with mission control, they removed the bag and watched as flight controllers orchestrated a successful docking.

The achievements of the *Kvant* module have political as well as scientific significance. The experiments carried out in *Kvant* were the first to attract significant Western participation in the Soviet space-station program. A major part of the science module is a high-energy astrophysics facility called Roentgen that is a collaborative endeavor between the Soviets and Western European scientific groups.

Four of the five telescopes in Roentgen were provided by prominent astrophysics research groups from England, West Germany, the Netherlands and Switzerland. The high-quality instruments, developed in anticipation of flights aboard satellites of NASA and the European Space Agency, enhance substantially the scientific value of *Kvant* and *Mir*. Indeed, the instruments in Roentgen were able to gather valuable data soon after entering operation: they made the first X-ray observations of the spectacular SN 1987a supernova.

But *Kvant* is more than a science module. It also carries "control-moment gyros" that help *Mir* to maneuver or maintain specific orientations. Previous Soviet stations used fuel-guzzling reaction-control jets for these purposes. The gyros save fuel, but unfortunately they also consume considerable electrical power. *Mir*'s average available power is now nearly 12 kilowatts, more than that available on earlier Soviet stations but still only half of that provided by the three fuel cells on the U.S. space shuttle.

Major construction will occur on *Mir* this year. The Soviets plan to add a 12.5-meter, 20,600-kilogram service module to *Mir* in April. The module will include new solar panels and a one-meter hatch to support a small, piloted vehicle used for working outside the station. Other equally large modules for technology studies and earth observations will be added later this year. And the Soviets apparently consider it important to extend the length of time their cosmonauts endure weightlessness; they are probably concerned with gathering data on its physiological effects.

SUBSCALE SPACE PLANE was photographed from a Royal Australian Air Force plane flying over a Soviet ship in the Indian Ocean. The people on the ship's deck were retrieving the model from the water, where it had landed after a test flight. The actual size of the space plane is probably about two or three times the size of this model.

Recognizing that psychological and sociological factors can limit cosmonaut endurance and performance, the Soviets designed *Mir* to be a more comfortable home than its orbital predecessors were. There is less bulky equipment in the living area, and each of the two resident cosmonauts has a closet-size "cabin" with a sleeping bag and a small window. Video and audio tapes are shipped to the crew regularly, and the cosmonauts maintain extensive contact with their colleagues, physicians, families and friends on the ground.

The daily schedule, however, is demanding, and facilities are crowded. Vladimir A. Shatalov, head of cosmonaut training at Star City and a former cosmonaut himself, remarked that "cosmonauts' living conditions are more severe than, say, those on submarines." One can then appreciate the personal qualities of Vladimir G. Titov and Musa K. Manarov, the two cosmonauts who returned to the earth in December after spending a year on board *Mir*.

From the foregoing discussion it is clear that the Soviet Union has established an active, well-orchestrated space-station program using technology that by Western standards is not highly advanced. The Soviets have sustained a continuing human presence in space and have mastered the logistics of complex space operations. The flow of cargo and cosmonauts to and from *Mir* is extensive and impressive. The components of their program have evolved in capability; they are reliable, they are produced in quantity and they have stood as effective building blocks in an increasingly ambitious agenda.

Yet space-station activities do not stress Soviet ground facilities or launch capacity. Of the roughly 50 annual SL-4 and Proton launches, only eight to 12 are needed to support the space-station program. In contrast, the assembly in the mid-1990's of the substantially larger U.S. space station Freedom will require approximately 20 shuttle flights. And because there are no plans for automated deliveries of cargo or fuel, station operations will probably require five to seven shuttle flights per year after Freedom is built. Five to seven flights is a sizable fraction of the shuttle's projected 12 to 14 flights per year.

The design of *Mir*'s space-station transport system, with cosmonauts ferried to *Mir* in Soyuz capsules and cargo carried in automated spacecraft, also gives the Soviets flexibility in mission planning and operations and enables them to assess the significant problem of human safety independently. Such flexibility has not been incorporated in plans for the U.S. space station.

Soviet scientific missions are limited not by launch capacity but only by the number of high-quality payloads they can supply. If the Soviets continue to attract Western participation as they have in *Kvant, Mir* and its successors could become important research facilities. Indeed, the Soviet Union's space program has already become a means of promoting scientific, economic and political ties with Western nations. For some time now the Soviets have used the cosmonaut program to strengthen relations with Soviet-bloc countries. Cosmonauts from all the Eastern European countries and from Cuba, Vietnam, Mongolia, France, Syria and Afghanistan have flown with the Soviet program.

The Soviets are now inviting even wider foreign participation in their space program, and they have embarked on a remarkably aggressive marketing campaign. Soviet launch vehicles and launch services are available at a price that is low compared with Western rates. High-resolution satellite images of the earth's surface (excluding the Soviet Union and Eastern-bloc countries) can be bought without difficulty, and satellite telecommunication services have been offered for sale. The technical apparatus on *Mir* can even be rented.

So far the Soviet marketing effort has met with little success. It should not be taken lightly, however. The valuable commodities of the Soviet program are bound to be attractive to many scientists, companies and countries that have traditionally relied on the U.S. for access to space.

What can one expect from the Soviet space-station program in the future? First, it is evident that the Soviets intend to continue improving their ability to transport cosmonauts and cargo into space. The Soviet space shuttle is one example of this intention. Although it appears to be a departure from the Soviets' evolutionary approach, the shuttle will enhance current transportation capabilities, not replace existing ones. For example, the shuttle will provide a means to return large payloads to the earth, a capability the Soviets currently lack. The Soviet shuttle will be one element of their program—not the centerpiece.

A more forward-looking addition to Soviet capabilities is the small, reusable space plane. This vehicle has not yet been acknowledged by Soviet officials. Its existence was revealed several years ago, however, when a Royal Australian Air Force plane flying over the Indian Ocean photographed a scaled-down model being retrieved by a Soviet ship after a test flight.

Western analysts expect that the space plane would be launched by an SL-16 and carry two or three people and perhaps a very small cargo into low earth orbit; it would land on a standard runway. The plane could also be used to transport people between space facilities, to inspect satellites or as a rapid-response rescue vehicle.

France recently persuaded the European Space Agency to develop a similar vehicle called Hermes, which is scheduled to be completed in the late 1990's. The U.S. Air Force studied the concept of a space plane during the early 1960's in the so-called Dyna-Soar project. The project was terminated in 1965 because of cost and the lack of identified uses for the plane in the military space program. Currently the U.S. has no plans to pursue this type of vehicle.

Planning for the next-generation Soviet space station is reportedly already under way. If Salyut history is any guide, *Mir* will be used for perhaps four more years before it is replaced. Energia will undoubtedly play a major role in plans for future Soviet space stations. It will probably put in orbit the core of the next-generation Soviet station. And the Soviets have hinted that Energia itself may be just one stage in an even larger rocket.

It is likely that the Soviets will continue their drive to extend the length of time cosmonauts remain in space. Yet even a large space station in low earth orbit would not require cosmonauts to remain aloft as long as Titov and Manarov did. Soviet life scientists themselves have estimated that human performance is maximized if the tour of duty on board a space station is between four and six months.

Does this preoccupation presage a major program to send cosmonauts to Mars? Perhaps, but it will certainly not happen in the near future. It is true that the development of Energia and an understanding of human response to weightlessness are necessary in order to reach out to Mars; it is also true that Soviet space representatives have been talking about Mars more frequently. But an expedition to Mars is a huge undertaking, much more demanding than the Apollo program. It would take from 15 to 20 Energia launches to put the Mars vehicle and its fuel into low earth orbit for assembly, and a new-generation space station would be needed to house assembly workers and to store, integrate and test the vehicle and its component systems. Such an extraordinary task would require a generation of dedicated effort.

The Soviets have taken great strides in their cosmonaut program using modest technological tools. Now Energia, the space shuttle and the prospective space plane provide the Soviets with new tools that will make possible even more ambitious space activities in the next decade.

The success of the Soviet program appears all the more impressive when it is viewed against the background of recent problems in the U.S. Questions concerning U.S. national space policies, plans and even capabilities have come to the fore in the extended aftermath of the *Challenger* accident. While a major Soviet space station logs its third year in orbit, the goals for our own station remain uncertain and the plans for its operation are at least seven years from realization. While Americans continue to debate their space-station program, the Soviets inhabit a humble but operative complex and make plans for a more capable facility.

Can the U.S. manage its own, more sophisticated space technology well enough to mount an equally powerful program? Certainly it can, but not without the ingredients that characterize the Soviet program: long-range political and financial commitments that nurture long-range goals.

FURTHER READING

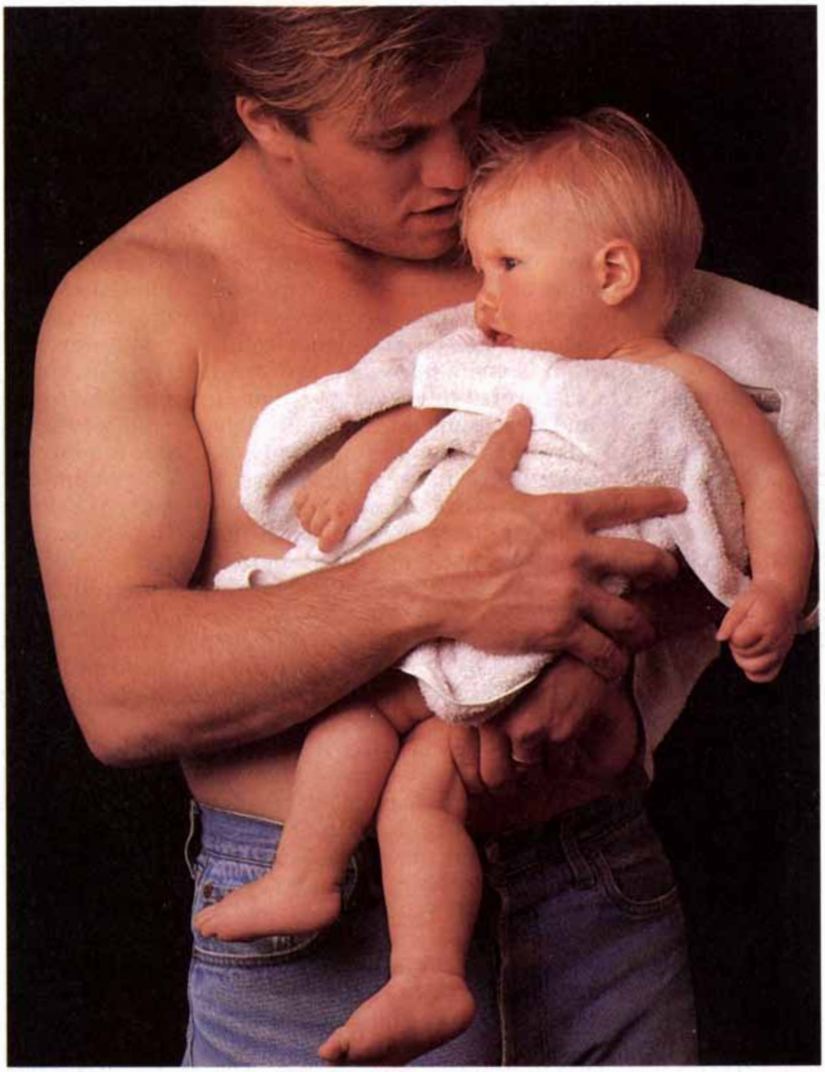
THE HEAVENS AND THE EARTH: A POLITICAL HISTORY OF THE SPACE AGE. Walter A. MacDougal. Basic Books, Inc., 1985.

SOVIET SPACE STATIONS AS ANALOGS. B. J. Bluth and Martha Helppie. System Engineering and Integration Group, Distributed Systems Branch, SSE, NASA; August, 1986.

THE SOVIET MANNED SPACE PROGRAM. Phillip Clark. Orion Books, 1988.

SOVIET SPACE PROGRAMS 1981–1987: PART I. U.S. Congress, Senate Committee on Commerce, Science, and Transportation. U.S. Government Printing Office, 1988.

THE SOVIET YEAR IN SPACE 1988. Nicholas L. Johnson. Teledyne Brown Engineering, Colorado Springs Office, 1250 Academy Loop, Suite 240, Colorado Springs, Colo. 80910; 1989.



John Elway. Cardmember since 1984.

Membership Has Its Privileges.℠

Don't leave home without it.®

Call 1-800-THE CARD to apply.

The Genes for Color Vision

Recently isolated, the genes encoding the color-detecting proteins of the human eye have yielded new clues about the evolution of normal color vision and the genetic bases of color blindness

by Jeremy Nathans

"Colors," said Leigh Hunt, a 19th-century poet, "are the smiles of Nature." Just how does an observer distinguish one smile from another? To a great extent the answer lies in the three classes of cone-shaped, color-sensing cells in the retina of the eye. Each class responds differently to light reflected from a colored object, depending on whether the cells have within them red, green or blue pigments: light-absorbing proteins particularly sensitive to wavelengths in either the long-wave (red), intermediate-wave (green) or short-wave (blue) region of the visible spectrum. The relative amounts of light absorbed by each class of cones are translated into electrical signals by retinal nerves and then transmitted to the brain, where the overall pattern evokes the sensation of a specific hue.

The role of the pigments in color discrimination has been known for decades, and yet their structures were not elucidated until recently. My colleagues and I have now identified the genes that code for the pigments, deciphered their structures and thereby deduced the amino acid sequences of the encoded proteins. This work, completed at the Stanford University School of Medicine, should soon lead to the isolation of the pigments themselves and a detailed examination of how they function. Our studies have also provided clues to the evolution of normal color perception and the variant color vision that is often called color blindness. (Actually the term is a misnomer: few people perceive no color at all.)

JEREMY NATHANS is assistant professor in the department of molecular biology and genetics and the department of neuroscience at the Howard Hughes Medical Institute of the Johns Hopkins University School of Medicine. After obtaining a bachelor's degree from the Massachusetts Institute of Technology and an M.D. and a Ph.D. from Stanford University, he spent a year as a postdoctoral fellow at Genentech. Nathans moved to Johns Hopkins in 1987.

The new findings add detail to a picture of color vision pieced together over the past several centuries. Isaac Newton made the first major contribution some 300 years ago when he discovered the color spectrum. He found that sunlight, or white light, decomposes into a continuous sequence of colors when it is refracted by a glass prism. He also recognized that light at each angle of refraction has a characteristic color, ranging from red for the least-refracted rays through orange, yellow, green, blue and violet for rays that are refracted progressively more. Today each angle of refraction, and hence each pure color, is known to correspond to light of a distinct wavelength.

Newton further observed that the human eye often cannot distinguish between colors formed by quite different combinations of lights. He found, for example, that certain pairs of lights with different angles of refraction, such as red and green, could be mixed to produce a color sensation indistinguishable from the sensation produced by a pure third light—yellow in this case—whose angle of refraction was intermediate between those of the original lights.

By the late 18th century workers had extended Newton's observations and learned that color vision is trichromatic. This means most colors can be matched by a mixture of three primary lights; in all other instances a match can be achieved by mixing two of the primaries as before but adding the third primary to the given color. A variety of monochromatic, or pure, lights can act as primaries, but all sets of primaries consist of one long-wave, one intermediate-wave and one short-wave light; when the three primaries are mixed in equal parts, they produce the sensation of white. By convention, red, green and blue lights are today considered to be "the" primaries.

The English physician and physicist Thomas Young suggested in 1802 that trichromacy is a reflection of human physiology. He proposed that the colors one sees are determined by the relative extents of excitation of three types of sensors. "As it is almost impossible to conceive each sensitive point of the retina to contain an infinite number of particles, each capable of vibrating in perfect unison with every possible undulation," he noted, "it becomes necessary to suppose the number limited, for instance, to the three primary colors." It took time, but Young was eventually shown to be correct. The three classes of sensors—the cones—are now known to have overlapping but distinct sensitivities to light. For instance, the red and green receptors both absorb orange light, but the red receptor does so more efficiently.

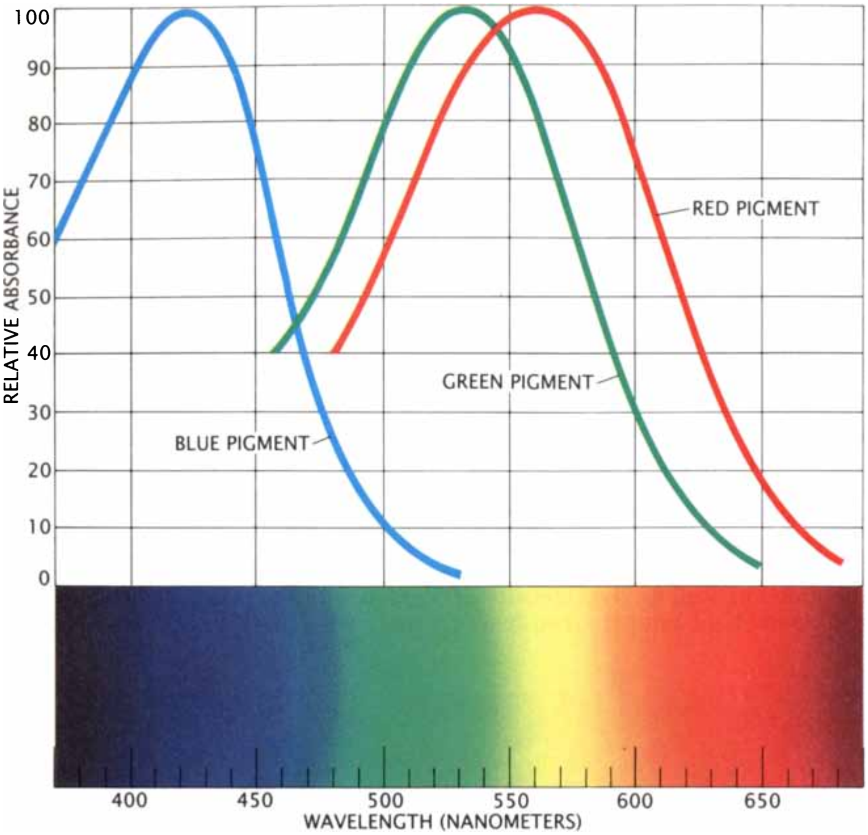
Shortly before Young put forward his theory, a contemporary, John Dalton (the father of atomic theory), helped to arouse interest in the study of variant color vision. Such work has paralleled and greatly informed the study of normal color vision. In Dalton's first paper to the Manchester Literary and Philosophical Society, published in 1794, he reported that he did not see colors as others saw them. "That part of the image which others call red appears to me little more than a shade or defect of light," he said. He also added that orange, yellow and green appeared to him as "what I should call different shades of yellow." Today deficiencies in the ability to discriminate between colors in the red-to-green region of the spectrum ("red-green discrimination"), which are found in roughly 8 percent of Caucasian males and 1 percent of females, are sometimes referred to as Daltonism. Defects in the ability to distinguish among colors in the blue part of the spectrum also occur but are rare; I shall not discuss them in detail.

In the mid-19th century the Scottish physicist James Clerk Maxwell identified two types of Daltonism by displaying various colors to his subjects and then systematically examining the colors they could not distinguish [see "Visual Pigments and Color Blindness," by W. A. H. Rushton; SCIENTIFIC AMERICAN, March, 1975]. Following Young's three-receptor theory, Maxwell estimated the light sensitivities of the three receptors and divided his color-variant subjects according to whether they confused colors that equally excited either the red and blue or the green and blue receptors; presumably subjects with normal color vision distinguished such colors by differences in the excitation levels of their green and red receptors respectively. Maxwell correctly inferred that one group of color-variant subjects was missing the green receptor and the other group was missing the red. I and others refer to these groups as green$^-$ (green-minus) or red$^-$ (red-minus) dichromats.

Later in the 19th century the English mathematician and physicist John William Strutt, better known as Lord Rayleigh, introduced the anomaloscope, a device that even to this day is the centerpiece of color-vision testing. The anomaloscope projects three different monochromatic lights onto a screen. In studies of color discrimination in the red-to-green part of the spectrum a deep red and a green light are projected onto half of the screen so that they mix, and a yellow light is projected separately onto the other half; subjects adjust the ratio of the red light to the green as well as the intensity of the yellow light until the two halves of the screen appear to be matched.

The design of the anomaloscope relies on the fact that people with normal color vision have two classes of color detectors—the red and the green—operating at the red-to-green end of the color spectrum. (The blue sensors are not affected by the test lights.) Subjects with normal color vision perceive a match when the red and the green sensors each absorb an equal amount of light (now quantified as an equal number of photons per second) from both sides of the screen. They match the yellow with a highly reproducible ratio of red light to green light [*see illustration on page 45*]. In contrast, red$^-$ and green$^-$ dichromats can match the yellow light with either the red or the green light alone; they can also match any ratio of red and green with the yellow light. They can make such matches because all three

RAINBOW emerges when sunlight passing through droplets of water is refracted so that it divides into its constituent colors. Isaac Newton's analysis of this and a similar phenomenon—the splitting of sunlight into a spectrum of color when it passes through a glass prism—led to centuries of research into color vision. It is now known that the color one sees depends on the relative activation of three classes of visual pigments, or light-absorbing proteins, in the retina. The cone-shaped cells housing the pigments translate the absorbed light into electrical signals and transmit them to the brain for interpretation.



SPECTRAL-SENSITIVITY CURVES show the sensitivity of the three color pigments to the visible spectrum of light. The curves were plotted on the basis of data obtained by James K. Bowmaker of the University of London and H. J. A. Dartnall of the University of Sussex. What is called the blue pigment is particularly responsive to the short-wavelength region of the spectrum; the green and red pigments are more sensitive to intermediate and long wavelengths. The pigments themselves have still not been isolated; hence Bowmaker and Dartnall determined their sensitivities by measuring the light absorbed by individual cones that were obtained from cadavers.

lights are detected by a single class of receptors; by adjusting the intensity of a single light, a dichromat can always equalize the number of photons captured from the two sides of the screen.

Such differences between normal subjects and red⁻ and green⁻ dichromats enabled Rayleigh to readily pick out the dichromats with his new device. He also identified two additional groups of color-variant subjects by testing his friends and relatives. Like normal subjects (and unlike the dichromats), these variant individuals required some mixture of red and green lights to match the yellow, but they chose unusual ratios. One group required more green and less red; the other group required the reverse. Rayleigh concluded that these subjects, now known as green-anomalous and red-anomalous trichromats, had green or red receptors with atypical spectral sensitivities.

By the mid-20th century such psychophysical studies, which depended on the judgment of subjects, strongly supported Young's three-receptor theory, and other studies had suggested that the cone cells of the retina served as those receptors. Finding direct proof for these ideas remained technically challenging, however. One major obstacle was that the cones were difficult to isolate. They are intermingled throughout the retina with the much more numerous rod cells, photoreceptors responsible for black-and-white vision in dim light.

Nevertheless, ingenuity won out. In the 1960's Paul K. Brown and George Wald of Harvard University and Edward F. MacNichol, Jr., William H. Dobelle and William B. Marks of Johns Hopkins University built microspectrophotometers capable of determining the absorbance of a single photoreceptor cell. The device passes a variable-wavelength light (one whose wavelength is changeable) through the color-detecting region of a cone and passes an identical beam through a different region, testing wavelengths across the spectrum; the difference in the intensity of the emerging beams at a specific wavelength is a measure of the absorbance of the color-detecting region at that wavelength. The studies employing the device demonstrated that cones taken at autopsy from human retinas did indeed exhibit three distinct absorbance spectra. The observed spectra agreed well with the sensitivities predicted by psychophysical studies.

A plot of the relative fraction of photons absorbed per second by each class of cones against the wavelengths of the visible spectrum yields three bell-shaped curves. The blue cones absorb wavelengths ranging from 370 nanometers (billionths of a meter) to 530 nanometers and are most sensitive to wavelengths of 420 nanometers. Both the green and the red cones are active across most of the spectrum but are particularly sensitive to wavelengths between about 450 and 620 nanometers. The green cones are most efficient at 535 nanometers, the red cones at 565 nanometers.

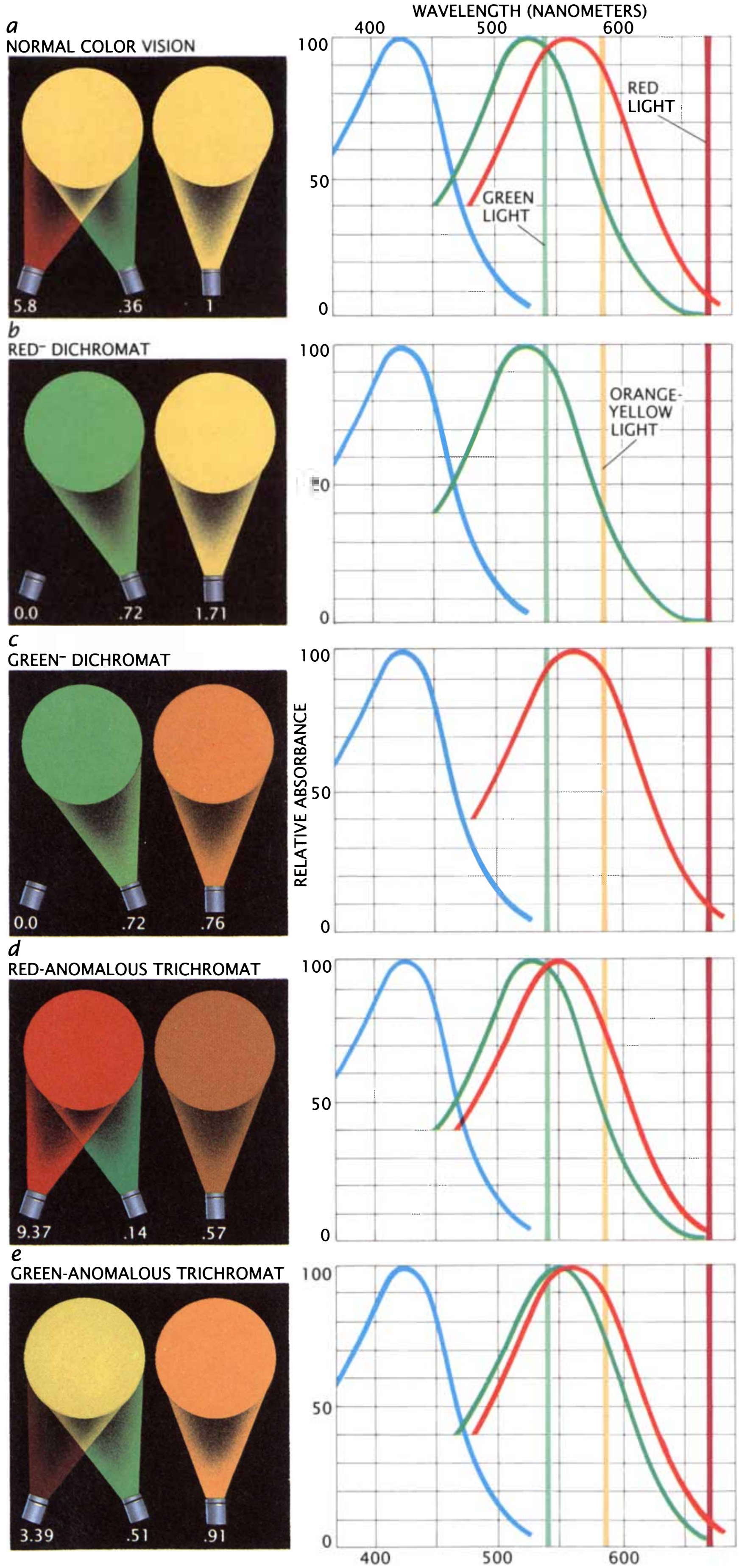
Beginning in the 1970's new evidence that dichromats lack one or another class of receptors emerged. William A. H. Rushton of the University of Cambridge directed a variable-wavelength light into the eyes of dichromats and measured the light reflected from—and hence not absorbed by—the retina; he thereby demonstrated that specific wavelengths are not absorbed normally by dichromats. More recently, James K. Bowmaker of the University of London, John D. Mollon of Cambridge and H. J. A. Dartnall of the University of Sussex showed with a microspectrophotometer that a retina obtained from a green⁻ dichromat did not have the green class of cones.

New evidence also yielded clues to the basis of anomalous trichromacy. By means of psychophysical techniques Rushton and, separately, Thomas P. Piantanida and Harry G. Sperling, who were then at the University of Texas, demonstrated that the spectral-sensitivity curves of anomalous red and green receptors lie in the interval between the normal red and green absorbance curves.

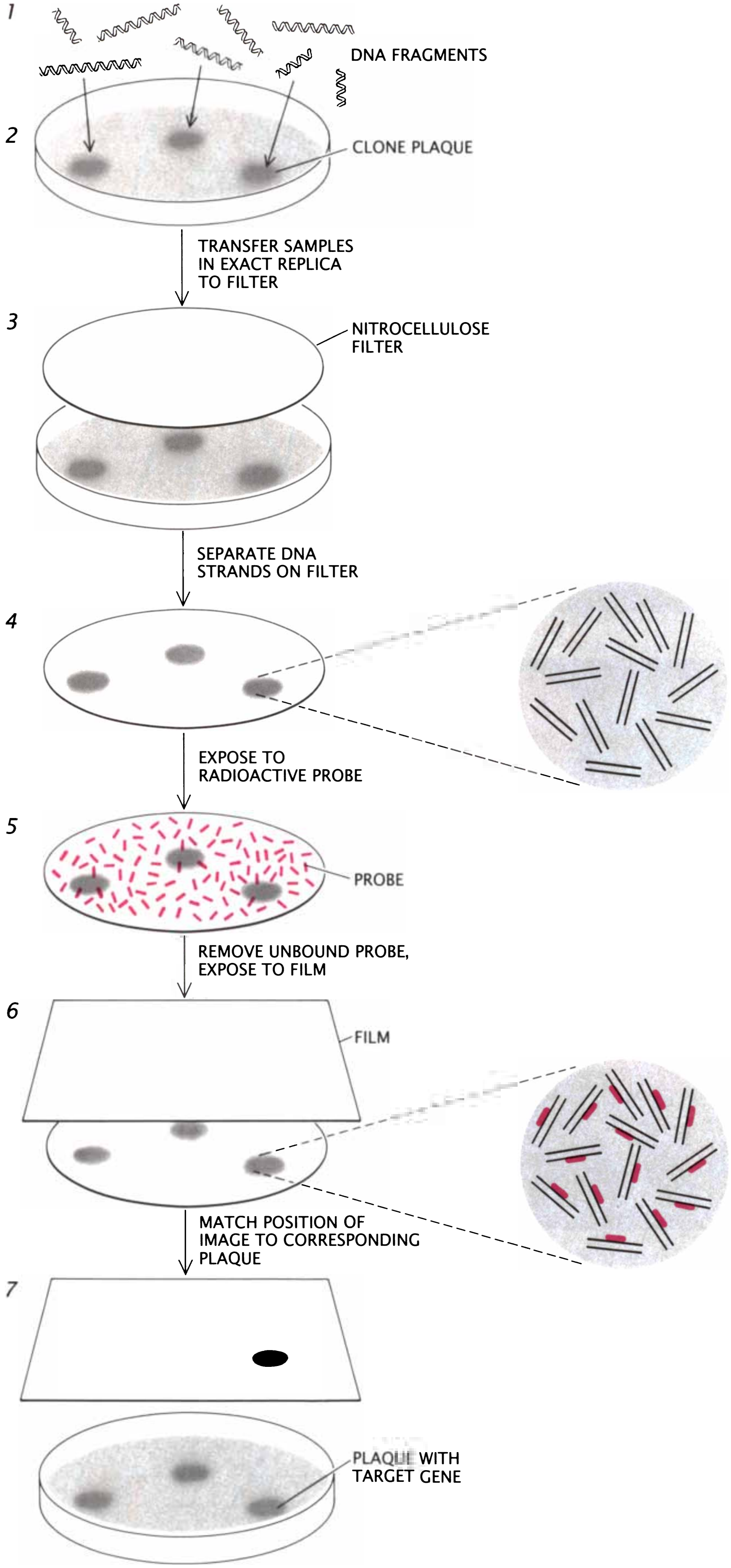
With the existence of three distinct classes of cones firmly established, in the early 1980's my colleague David S. Hogness and I turned our attention to the genetic bases of normal and abnormal color vision. In addition to aiding in the effort to isolate the pigment proteins, we hoped to supplement the findings of classical genetic studies that traced the inheritance of aberrant color vision within families.

Such studies built on the old observation that deficiencies in red-green discrimination are commoner in males than in females. Analysis of this pattern indicated that genes on the X (sex) chromosome are responsible for the variation. Males will have variant red-green discrimination if their single X chromosome (inherited from the mother) carries the trait; females will be affected only if they receive a variant X chromosome from both parents. The studies also indicated that variations in blue sensitivity are rooted in a gene on some nonsex chromosome. We planned to test the most straightforward explanation for these inheritance patterns, namely that color-vision variations result from inherited alterations in the genes encoding the cone pigments. Presumably mutations in such genes could result in the loss of a functional pigment or in the production of a pigment with an abnormal absorbance spectrum.

Our approach involved isolating the genes that code for the color pigments and comparing their structures in people with normal and variant color vision. Frequently one isolates genes by determining the amino acid sequences of the proteins they encode and parlaying that information into clues about the structure



COLOR MATCHES (*left*) made in a test of red-green color discrimination are depicted next to the subjects' spectral-sensitivity curves (*right*). The testing device, a Rayleigh anomaloscope, superposes a red and a green light on one half of a white screen and projects an orange-yellow light onto the other half. Subjects adjust the ratio of red light to green and the intensity of the yellow light until the two sides of the screen appear to match—that is, until the light absorbed by each type of pigment from one side of the screen is equal to the amount absorbed from the other side. The numbers below the lights show the relative intensities. Normal subjects (*a*) require both the red and the green light to match the yellow, choosing a high intensity of red and a low intensity of green. (The red and green pigments both absorb the red light less efficiently and the green light more efficiently than they do the yellow.) People who lack the red (*b*) or the green (*c*) pigment are identified by their ability to match the yellow light with the red or the green alone. (Only the green matches are shown.) Subjects whose red (*d*) or green (*e*) pigments have abnormal spectral sensitivities require both red light and green light to match the yellow but, compared with normal subjects, choose an excess of red or green respectively.



of the genes. Because virtually nothing was known about the structure of the color-pigment proteins when we began our work, we settled on a less direct strategy.

We started with the assumption that the cone pigments and the rod pigment, rhodopsin, all evolved from a common ancestral visual pigment and that as a result the present-day genes probably have some similar sequences of nucleotide bases: the four different chemical units whose sequence along the DNA helix encodes information. If we knew the structure of the rhodopsin gene, we reasoned, we could learn something about the structure of the cone-pigment genes. At the time neither human rhodopsin nor its gene had been isolated, but investigators had successfully isolated bovine rhodopsin from the retinas of cattle. Even better, Yuri A. Ovchinnikov and his colleagues at the M. M. Shemyakin Institute of Bioorganic Chemistry in Moscow and Paul A. Hargrave and his co-workers at Southern Illinois University had deciphered the amino acid sequence of the protein. That information would be our springboard for identifying the cone-pigment genes.

We planned to isolate the bovine rhodopsin gene and then use it as a probe to identify the human rhodopsin gene and the cone-pigment genes. The plan relied on a technique known as DNA hybridization, which exploits the fact that a single strand of DNA

DNA HYBRIDIZATION enabled the author and his colleagues to identify the genes for human rhodopsin—the pigment responsible for (colorless) vision in dim light—and its close relatives, the cone pigments. They chemically cut double-strand DNA (*1*), separated the fragments by size and cloned them in bacterial viruses grown on a lawn of bacteria. Each virus multiplied to form a plaque, a zone of viral multiplication (*2*). Nitrocellulose filter paper was placed over the dishes so that a sample from every plaque adhered (*3*). The DNA on the filter was chemically separated into single strands (*4*) and exposed to a radioactively labeled probe (*red*): a single-strand DNA from the bovine rhodopsin gene (*5*). If the probe was structurally similar to the genes for the human pigments, as it was expected to be, it would hybridize with (bind to) the pigment genes. To determine whether the probe did in fact bind, the filters were covered with photographic film (*6*). Dark spots appeared at binding sites (*7*), indicating the radioactive probe had bound. The workers then identified the clones containing the pigment genes by aligning the spots on the film with the corresponding plaques.

will form a stable double helix with a second strand if the sequence of nucleotide bases along one strand is complementary to the sequence along the other strand. In particular, the base adenine always pairs with thymine, and guanine always pairs with cytosine.

The hybridization technique has many steps, but in essence an investigator chemically cuts up the double strand of DNA to be probed and produces multiple copies of each fragment by cloning it. A sample from each clone is split into single-strand DNA, and a radioactively labeled probe, another piece of single-strand DNA, thought to be complementary to the gene one hopes to identify is added to the samples. If all goes well, the probe binds stably to its complement, thereby picking out the target gene [*see illustration on opposite page*].

Normally we would set about developing a probe that would accomplish our first aim—identifying the bovine rhodopsin gene—by generating a list of all possible base sequences that could give rise to the known amino acid sequence of rhodopsin. Then we would test a variety of probes constructed on the basis of that list. Fortunately we were spared much of the time-consuming process, because a probe for identifying the bovine rhodopsin gene was at hand. H. Gobind Khorana, Daniel D. Oprian and Arnold C. Satterthwait of the Massachusetts Institute of Technology and Meredithe L. Applebury and Wolfgang Baehr of Purdue University had already identified a DNA sequence that bound efficiently to messenger-RNA molecules encoding bovine rhodopsin. (Messenger RNA is the single-strand molecule that carries information from DNA in the nucleus to the cytoplasm, where it directs production of the encoded protein. RNA is virtually identical with the coding strand of the DNA from which it is transcribed.) Based on the DNA sequence of Khorana and Applebury and their colleagues, we synthesized a probe and used it to identify the bovine rhodopsin gene.

In the second stage of our plan we enlisted a strand of the newly identified bovine rhodopsin gene as a hybridization probe to search for the human rhodopsin and color-pigment genes. The probe bound strongly to only one segment of human DNA, which subsequent tests identified as the gene encoding human rhodopsin. Our probe also bound, although not as strongly, to three other DNA segments. When we determined their nucleotide sequences, we found the segments had coding regions that were homologous to those of the human and bovine rhodopsin genes. Analyses of the amino acid sequences of the encoded proteins showed that the molecules were also similar to one another: some 40 percent of each protein chain was identical with the sequence of rhodopsin.

We naturally suspected that the DNA segments bound by our probe were the three cone-pigment genes, but we wanted further evidence. For instance, we hoped that messenger RNA corresponding to the probe-bound sequences would be found in the retina of the eye, the only place where visual pigments, and hence visual-pigment RNA's, are produced. Sure enough, the retina did yield RNA corresponding to the probe-bound DNA.

To evaluate whether our findings were consistent with those of classical genetic studies, which would provide more evidence that we had found the color-pigment genes, we determined where on the chromosomes the DNA segments pinpointed by our probe lie. In collaboration with Thomas B. Shows and Roger L. Eddy of the Rosewell Park Memorial Institute in Buffalo, N.Y., we found that two of the three weakly hybridizing genes reside in exactly the region of the X chromosome where classical analyses had placed the source of variant red-green discrimination. We therefore concluded that these genes encode the red and green pigments, and later studies we did with Piantanida confirmed this belief. The third gene, now known to encode the blue pigment, came from chromosome 7, a finding that is consistent with the notion that variant blue color vision is determined by a nonsex chromosome.

The significant homology between the rhodopsin gene and the three cone-pigment genes suggested that all four genes had indeed evolved from the same ancestor. The available evidence supported the notion that at some early stage a primordial gene had given rise to three others: the rhodopsin gene, the blue-pigment gene and a third gene that encoded a pigment sensitive to light in the red-to-green part of the visible spectrum. This third gene recently duplicated, yielding a red- and a green-pigment gene.

We think the red- and green-pigment genes are the product of fairly recent duplication, because they have a strikingly high degree of homology: a full 98 percent of their DNA is identical, suggesting it has had little time to change. The idea that the event took place not long ago, at least in evolutionary terms, is supported by the findings of Gerald H. Jacobs of the University of California at Santa Barbara, who worked with Bowmaker and Mollon. They have shown that New World (South American) monkeys have only a single visual-pigment gene on the X chromosome. In contrast, Old World (African) monkeys, which are more closely related to human beings, appear to have two visual-pigment genes on that chromosome. The addition of the second X-chromosome gene must have occurred sometime after the separation of South America and Africa, and hence of the gene pools of the New and Old World monkeys, some 40 million years ago.

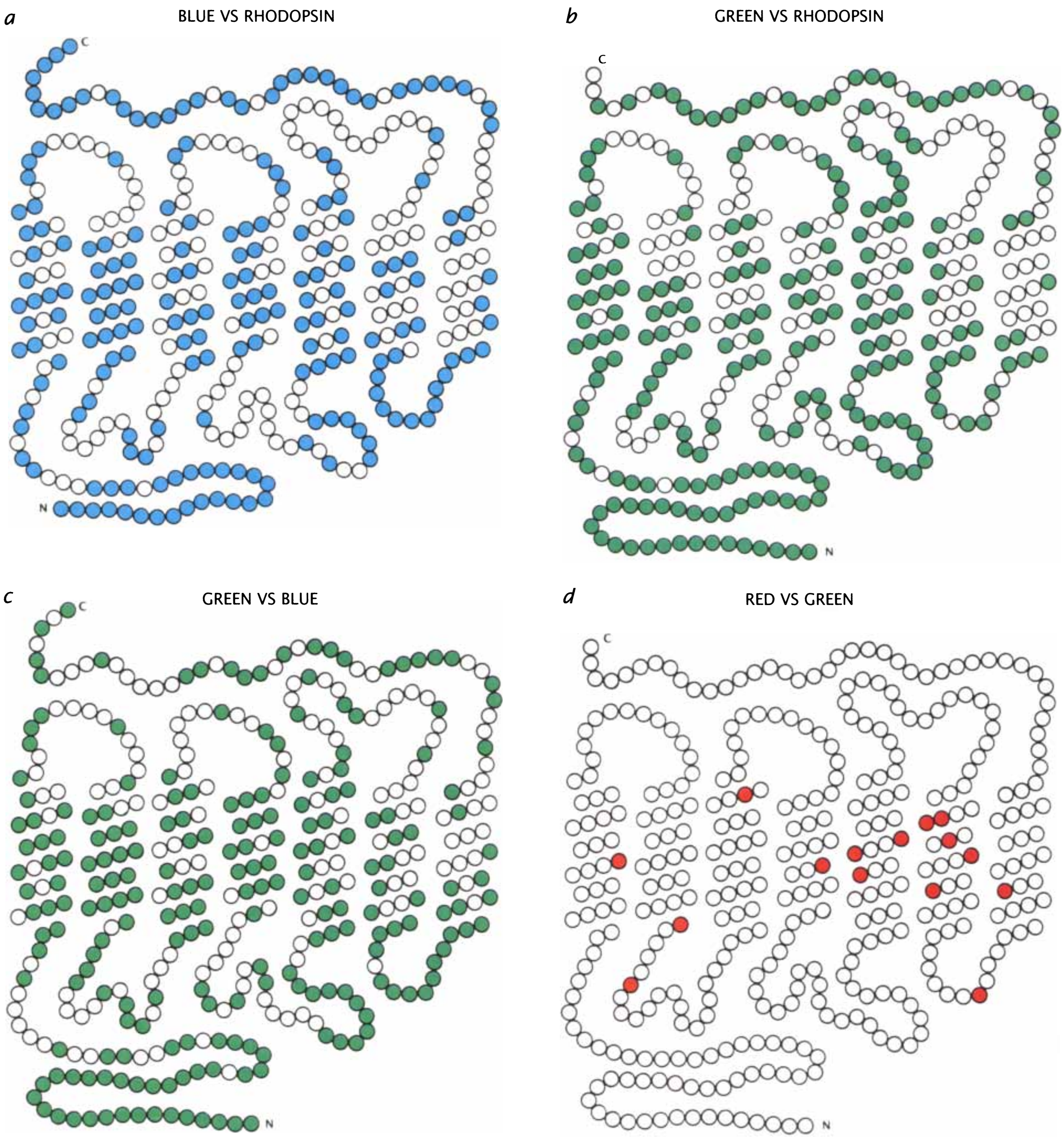
The discoveries described so far were not entirely unexpected, but another finding was. When we studied the visual-pigment genes from the X chromosome of 17 of our male colleagues, all of whom had normal color vision, we found that the red-pigment gene was always present in a single copy, but the other gene—the one encoding the green pigment—was present in one, two or three copies. The existence of multiple copies was surprising, because one green-pigment gene is presumably sufficient for normal color vision.

Recent experiments suggest that the visual-pigment genes lie in a head-to-tail array on the X chromosome and that the tandem arrangement accounts for the variability in gene number. Tandem genes that are similar have a tendency to undergo changes in copy number in the course of meiosis, the process of cell division that gives rise to sperm and eggs. Cells carry somewhat different versions of each chromosome; during meiosis the matching chromosomes pair up and recombine, or swap segments. Ordinarily the exchange is equal, so that each chromosome neither gains nor loses genes. Occasionally, however, two segments that are highly homologous can recombine erroneously, undergoing what is called unequal homologous recombination. Either one chromosome gains one or more copies of an existing gene at the expense of the other chromosome or the two chromosomes swap material from related but different genes. The chromosomes that result are then passed on in sperm or eggs.

It is quite easy to imagine how unequal homologous recombination might have given rise to the varied configuration of green-pigment genes in people with normal color vision. Consider two matching chromosomes, each carrying one red-pigment gene next to two green-pigment genes. If a green-pigment gene on one of the chromosomes crossed over to the other chromosome during meiosis, one daughter cell would carry a chromosome that had one red- and one green-pigment gene, but the other would have one red-pigment gene and three green-pigment genes—precisely the kind of variation seen in our subjects with normal color vision. There is only a single red-pigment gene in every case because that gene lies at the very edge of the array of color-pigment genes. A gene in that position is highly unlikely to be duplicated (or deleted) by homologous recombination.

Unequal homologous recombination appears to be responsible not only for the duplication of green-pigment genes in people with normal color vision but also for the great majority of deficiencies in red-green discrimination. In collaboration with Piantanida we studied DNA from 25 men who according to Rayleigh anomaloscope tests had variant red-green discrimination. All but one of the subjects had abnormal configurations of red- and green-pigment genes as a consequence of unequal homologous recombination.

Which configurations of genes yield



STRUCTURAL COMPARISONS of the four visual pigments on the basis of their nucleotide sequences indicate that rhodopsin and the color pigments all have similar amino acid sequences. (Each colored dot represents an amino acid difference.) The red and the green pigments are most alike; in fact, they are almost identical with one another (*d*). When the molecules being compared have different lengths, the longer molecule is depicted and the unshared "extension" is colored in.

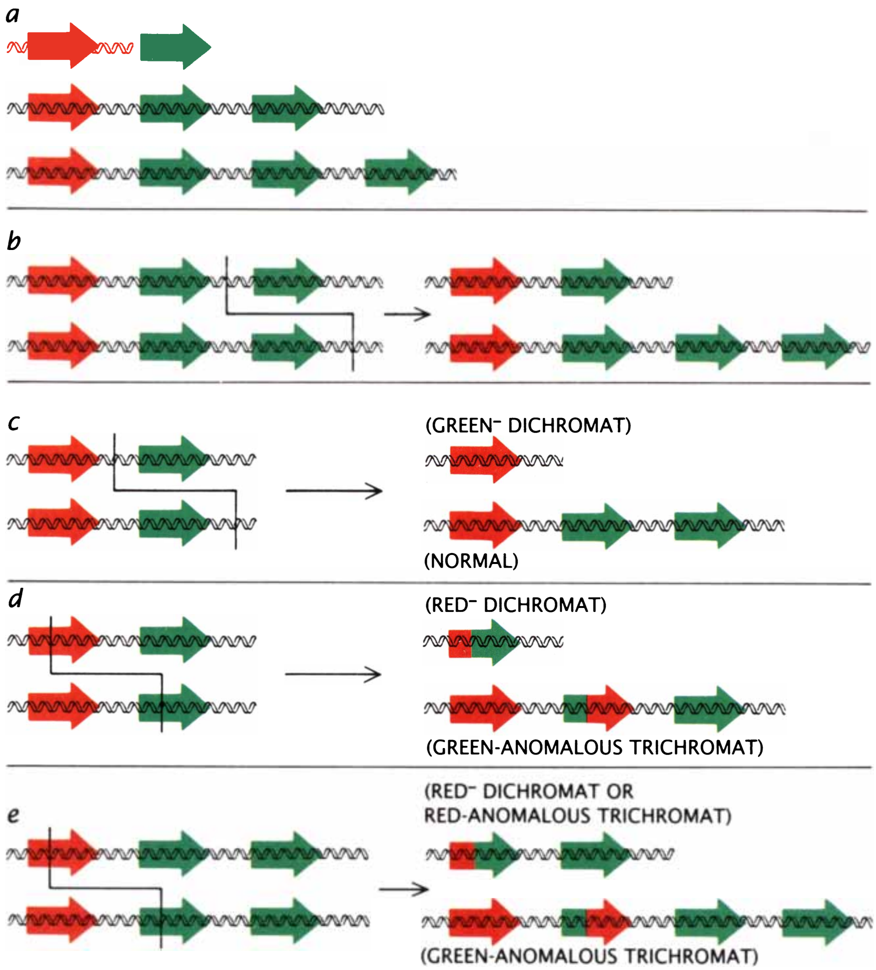
red^{-} or green^{-} dichromats and which produce anomalous trichromats? We found that most of the green^{-} dichromats (those who apparently lacked the green receptor) had simply lost all green-pigment genes. In some green^{-} dichromats, however, the green-pigment gene had been replaced by a hybrid gene: the DNA sequences near the start of the gene were derived from a green-pigment gene and the remaining sequences were derived from a red-pigment gene. Apparently the chromosome with the hybrid gene resulted from recombination in which part of the normal green-pigment gene changed places with part of a red-pigment gene.

Why did the hybrid gene not result in a functioning green receptor? It seems likely that the DNA sequences at the start of either normal or hybrid genes determine the cell type in which the gene will be active, and that the more distal sequences determine the type of pigment produced. This would result in the production of red pigment in cells that would normally become green cones, allowing them to function only as red receptors.

Among men whose Rayleigh anomaloscope tests indicated an absence of the red receptor (red^{-} dichromats), the great majority did not lack red-pigment sequences entirely. Instead their single red-pigment gene had been replaced by a hybrid in which only the initial DNA sequences were from a red-pigment gene. We surmise that this hybrid gene results in the production of a green pigment in cells that would normally have become red cones, in effect endowing people who carry the gene with all green cones and no red ones.

All the subjects with anomalous trichromacy had at least one hybrid gene in addition to some or all of the normal visual-pigment genes. We suppose that in these individuals the hybrid genes encode proteins that have anomalous spectral sensitivities. Our findings suggest that a variety of anomalous pigments are possible and that the particular spectral sensitivities of these pigments are determined by the exact point of crossing-over within the hybrid gene. The larger the fraction of the hybrid derived from the green-pigment gene is, the more greenlike the encoded pigment will be; likewise, if a large fraction of the hybrid is derived from the red-pigment gene, the encoded pigment will be more redlike. If we are correct, our data would account well for the psychophysical observation that anomalous receptors most efficiently absorb wavelengths lying in the interval between the wavelengths that are most efficiently absorbed by normal red and green cones.

Exciting questions concerning the role of cones and pigments in color vision remain. What is it about the visual pigments that gives them their distinctive absorbance spectra? How does each photoreceptor cell determine which visual pigments to produce? How are connections between photoreceptor cells and high-order neurons formed during development? For those seeking answers to these and related questions, inherited variations in human color vision are a gift, offering a unique window to the inner workings of the eye.



EXCHANGE OF GENETIC MATERIAL between normal X chromosomes, which bear the red-pigment and green-pigment genes (*colored arrows*), can give rise to either normal or abnormal red-green color discrimination. X chromosomes derived from normal individuals have one normal red-pigment gene and one, two or three green-pigment genes (*a*). Such normal variation can readily arise when a chromosome with one red- and two green-pigment genes loses one of the green-pigment genes to its mate (*b*). Dichromacy and anomalous trichromacy commonly arise from genetic exchanges that result in the loss of a pigment gene (*c*) or the creation of a hybrid gene derived in part from a red- and in part from a green-pigment gene (*d, e*). The light sensitivity of a hybrid pigment depends on the point of crossing-over within the hybrid gene.

FURTHER READING

THREE-PIGMENT COLOR VISION. Edward F. MacNichol, Jr., in *Scientific American*, Vol. 211, No. 6, pages 48–56; December, 1964.

THE MOLECULAR BASIS OF VISUAL EXCITATION. George Wald in *Nature*, Vol. 219, No. 5156, pages 800–807; August 24, 1968.

MOLECULAR GENETICS OF HUMAN COLOR VISION: THE GENES ENCODING BLUE, GREEN, AND RED PIGMENTS. Jeremy Nathans, Darcy Thomas and David S. Hogness in *Science*, Vol. 232, No. 4747, pages 193–202; April 11, 1986.

MOLECULAR GENETICS OF INHERITED VARIATION IN HUMAN COLOR VISION. Jeremy Nathans, Thomas P. Piantanida, Roger L. Eddy, Thomas B. Shows and David S. Hogness in *Science*, Vol. 232, No. 4747, pages 203–210; April 11, 1986.

Giant Ocean Cataracts

Undersea cataracts that descend farther than any waterfall and carry more water than any river play a crucial role in maintaining the chemistry and climate of the deep ocean

by John A. Whitehead

Trivia enthusiasts know that Angel Falls in Venezuela, with a height just short of a kilometer, is the world's tallest waterfall, and that the Guaíra Falls along the Brazil-Paraguay border has the largest average flow rate: about 13,000 cubic meters per second. Trivia enthusiasts have not peered below the Denmark Strait. There an immense cascade of water—a giant ocean cataract—carries five million cubic meters of water per second through a descent of 3.5 kilometers, dwarfing Angel Falls in height and Guaíra Falls in flow rate. Even the mighty Amazon River, which dumps 200,000 cubic meters of water into the Atlantic every second, pales beside the Denmark Strait cataract. And although even the largest waterfalls on land are trivial components of the earth's climatic balance, giant cataracts play a vital role in determining the temperature and salinity of the deep ocean.

Ocean cataracts have been seriously investigated by oceanographers only within the past 20 years. They are a direct result of the process of convection: the transfer of heat by the bulk motion of a fluid. One can imagine the ocean to be a shallow pan of water that is exposed to the sun at one end (the Tropics) and not at the other (the high latitudes) [*see illustration on page 52*]. The cold water near the pole is the densest and therefore sinks in convective currents to the bottom of the pan. From there it spreads toward the temperate latitudes, displacing the warmer water above it. The warmer water therefore begins to rise in a gentle upwelling that is thought to take place almost everywhere in the ocean. Because warmer layers prevent bulk upward motion of the colder bottom layers—just as a Los Angeles temperature inversion traps cold air under warm—the upwelling is extremely slow. At the same time the cold water is heated by contact with the warmer layers above.

The ocean is in thermal equilibrium, and so the heat flowing upward in this process must equal the heat flowing downward. But because convection is an extremely efficient mechanism for transferring heat, the downward convective currents do not have to be very large in cross-sectional area to balance the heat transferred by the oceanwide warming of the deep water. The narrow currents of sinking cold water are in fact the precursors to the ocean cataracts proper.

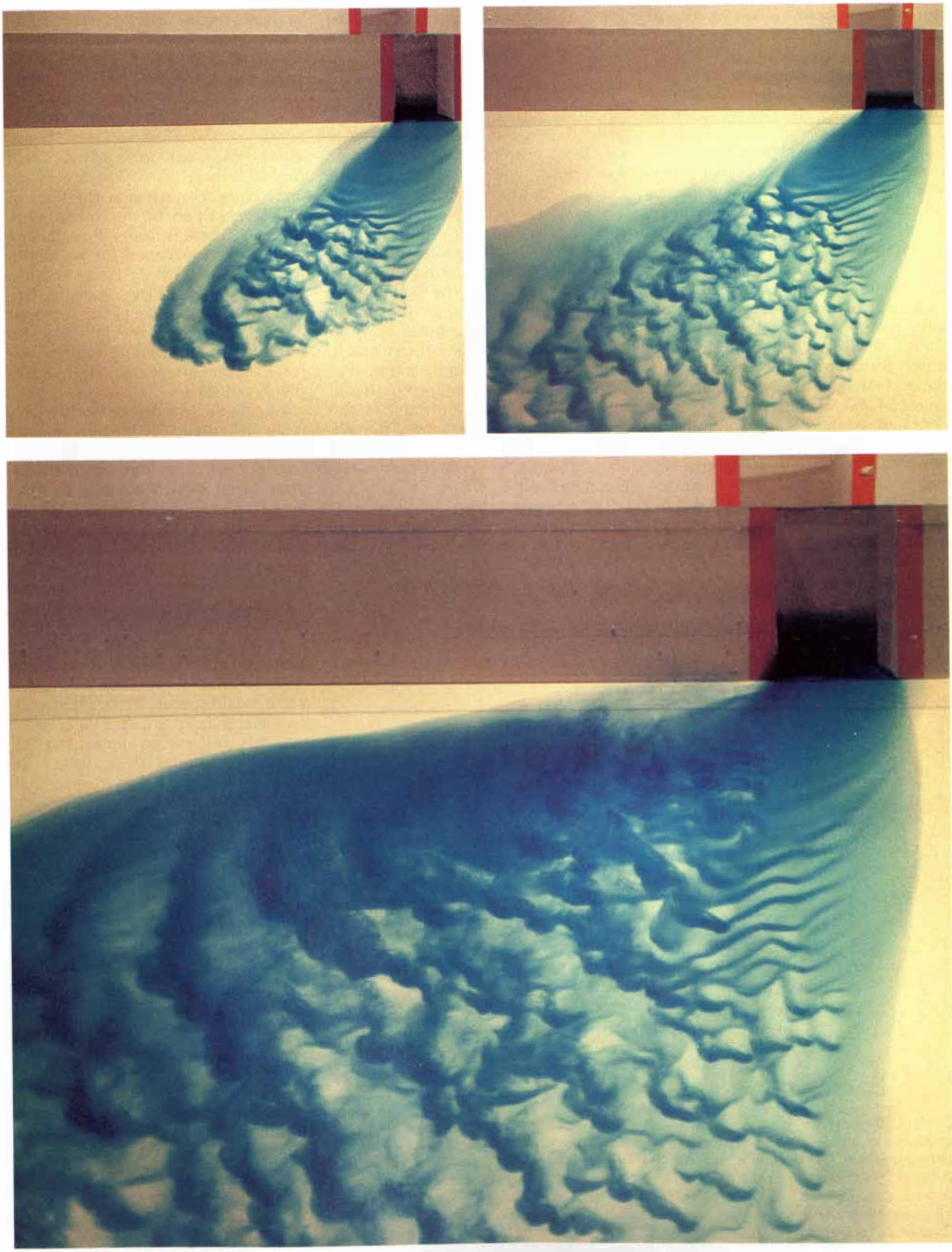
It is easy to estimate the amount of time over which the upwelling takes place. If the ocean pan has a volume of roughly 3×10^{17} cubic meters and is being fed by a stream of cold water at five million cubic meters per second, then to fill the pan with cold water would take about 2,000 years. If the pan is five kilometers deep, and no heat is transferred from the warm upper layers to the cold lower layers, the cold-water level would rise at a rate of from two to three meters per year. Although this is a crude estimate, it corresponds roughly to the rate of upwelling thought to take place in the real ocean.

Of course, the real ocean is not as simple as a pan of water; it contains topographical features that divide it into a series of large basins. Suppose that the ocean pan is divided into a polar and an equatorial basin by a hump, or sill. Cold water sinking at the pole can now fall into only the polar basin. To reach the lower latitudes the dense water must first rise above the sill; it does this by the gradual upwelling discussed above. During the process the water is heated by contact with the warmer layers above, but because the conductivity of water is very low, the heat transfer is minimal and the temperature change is not large.

Consequently the water temperature at sill depth in the polar basin is less than the temperature at sill depth in the equatorial basin. The polar water is therefore denser than the equatorial water, and so it spills over the sill, displacing the warmer water. I shall refer to the resulting current across the sill as an ocean cataract. As it flows from the polar basin into the equatorial basin, the cataract skims off warmer water encountered at or above sill depth and carries it downward to the bottom of the second basin. Moreover, as the cataract flows downward it often mixes turbulently with the surrounding downstream water. The net result of the skimming and mixing is that the water temperature at the bottom of the equatorial basin equals—or even exceeds—the water temperature at sill depth in the polar basin.

The study of real-world cataracts must take into account the fact that oceans contain many basins. Certain complicating factors that I shall discuss below, such as the Coriolis force and friction, also enter the picture, with the result that the paths of some cataracts are not oriented from north to south and do not end at the Equator. Nevertheless, the simple model accounts for most of the observations: cataracts of cold, dense water flow from the polar basins into the temperate-latitude basins. At each stage the water is warmed through skimming and mixing. Thus the bottom tempera-

JOHN A. WHITEHEAD is a senior scientist in the department of physical oceanography at the Woods Hole Oceanographic Institution. He received a B.S. at Tufts University and an M.S. and a Ph.D. from Yale University, the last in 1968. In addition he is a past holder of a Guggenheim fellowship and is a fellow of the American Physical Society. His research interests center on analytical and laboratory studies of fluid mechanics in oceans, atmospheres and planetary interiors. For relaxation Whitehead plays the trombone in the Falmouth, Mass., town band, which is conducted by his wife, Linda.



LABORATORY CATARACT is created by introducing a stream of salt water into a rotating tank of fresh water. The denser salt water, here colored blue, begins to sink. As it descends along the sloping bottom of the tank it develops waves and mixes turbulently with the surrounding water. The Coriolis force arising from the tank's rotation causes the cataract to veer to its right. A difference in salinity also drives the cataract that flows out of the Strait of Gibraltar into the North Atlantic.

ture of ocean basins increases toward the Equator.

Because the giant ocean cataracts occur at great depths in limited areas, they are not easy to study. Although oceanographers alluded to their presence as early as the 1870's, it was not until the 1960's that extensive investigation of the phenomenon became possible. The breakthrough came when vacuum tubes were replaced by transistors—leading to electronic equipment that could withstand being thrown overboard.

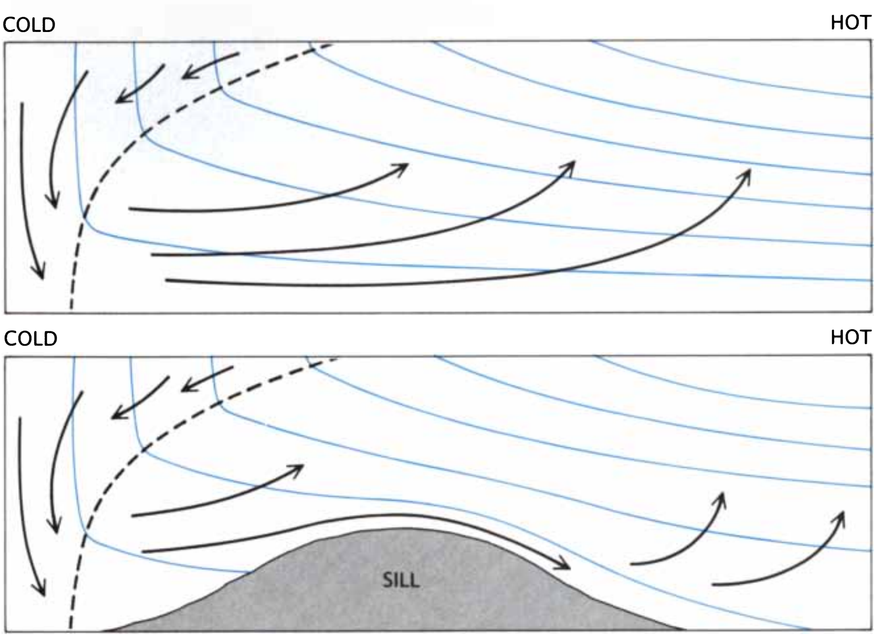
One way to find cataracts is to examine a north-south ocean slice that includes the suspected region of cold-water formation—the region where the cold water sinks [*see illustration on pages 54 and 55*]. One can then plot a series of isotherms, or contours of constant temperature. A cataract carries cold water downward over a sill, and so if a cataract is present, an isotherm just above sill depth will be roughly level on the upstream side of the sill but will descend sharply on the downstream side. Colder water, lying just below sill depth, will not flow over the sill; consequently the isotherms representing its temperature will be found in the upstream basin but not in the downstream one.

A good example of this pattern is provided by the Rio Grande cataract, which begins at about 20 degrees south latitude in the Atlantic Ocean and flows north. The Rio Grande Rise itself is the sill for this cataract; it lies at 4,000 meters below sea level. The isotherm for a potential temperature of zero degrees Celsius is slightly deeper (potential temperature is the actual temperature corrected for the effect of pressure in the deep ocean). Therefore the sill blocks this water, which originates in the Antarctic, from spilling into the more northern Brazilian Basin. Slightly warmer water, at .2 degree C, lies above the crest of the sill and is denser than the water to the north; this water flows over the sill, and the .2-degree isotherm descends about a kilometer to the bottom of the Brazilian Basin.

The one-degree and 1.4-degree isotherms reveal a second cataract farther to the north, where the Ceara Abyssal Plain separates the Brazilian Basin from the North Atlantic Basin. The plain, lying close to the Equator, prevents one-degree water in the Brazilian Basin from flowing into the North Atlantic Basin. On the other hand, the 1.4-degree water lies above the sill, which is thought to be at about four degrees north latitude. To the south the 1.4-degree isotherm is relatively level, but north of the sill it descends more than 2,000 meters and apparently terminates at the floor of the deepest part of the North Atlantic Basin. There it forms the Antarctic bottom water—named for its southern origins—of the North Atlantic. The termination of the 1.4-degree isotherm probably reflects the fact that during its descent the falling water mixes with surrounding warmer water; its temperature rises and the isotherm vanishes.

Tracing the two-degree isotherm reveals the Atlantic's third great cataract, the Denmark Strait overflow. The current that flows through the Denmark Strait (which actually lies between Greenland and Iceland) travels from north to south, opposite to the flow of Antarctic bottom water in the cataracts I have discussed above. Once the current crosses the sill, called the Greenland-Iceland Rise, it descends as a cataract some 200 meters deep and 200 kilometers wide; 1,000 kilometers downstream it reaches a depth of about 3,500 meters and forms the North Atlantic deep water. During its 3.5-kilometer descent the falling water may mix with warmer water, increasing its temperature above that of the Antarctic bottom water. In any case, the North Atlantic deep water is not as dense as the Antarctic bottom water and consequently forms a distinct layer above it.

Thermometers are not the only means by which oceanographers can probe cataracts. Radioactive isotopes dissolved in seawater also provide an effective way to trace the paths of these underwater currents. Tritium, a radioactive isotope of hydrogen with a half-life of 12.5 years, was produced at levels far above natural background levels by the atmospheric nuclear-bomb tests in the 1950's and early 1960's. Most of the tritium was released in the Northern Hemisphere and entered the ocean in the form of tritiated, or heavy, water. With the termination of aboveground testing in 1963 that source of tritium disappeared, and atmospheric levels of the isotope have declined steadily since. Thus a large volume of water that lay at the surface during the years of atmospheric testing was labeled with tritium.



CONVECTION drives most ocean cataracts, as is demonstrated in a pan of water when one end is heated and the other is cooled (*top*). The cold water at the "pole" sinks rapidly to the bottom of the pan (*arrows*) and spreads toward the "equator." There it encounters warmer water layers above. This temperature inversion prevents the cold water from rising rapidly; instead it is gradually heated by contact with the warmer layers and rises at a rate of about a meter per year in the ocean. A cataract is formed when a hump, or sill, is introduced into the pan (*bottom*). Water in the polar basin is colder and hence denser at the sill depth than water in the equatorial basin; the effect is manifested by the raising of the isotherms, or lines of constant temperature (*solid lines*), on the left-hand side. The polar water therefore flows over the sill to the bottom of the equatorial basin. This rapid descent, accompanied in some cases by turbulent mixing, is a model for a giant ocean cataract.

In 1972 the Geochemical Ocean Sections Study (GEOSECS) Operations Group measured tritium levels at various depths along a north-south slice of the Atlantic. The data showed that tritium from the atmospheric tests had been transported by the Denmark Strait cataract to a depth of roughly 3,500 meters, into the base of the North Atlantic deep water. The GEOSECS project found no tritium in the deeper layer of Antarctic bottom water, however. This indicates that the tritium deposited in the Southern Hemisphere has not had time to reach the Northern Hemisphere by way of the deep currents. At the ocean turnover rate estimated above, hundreds more years will elapse before tritium will appear in the Antarctic bottom water of the North Atlantic.

Tritium is generally difficult to detect in the Southern Hemisphere because the bomb tests were held primarily in the Northern Hemisphere. In contrast, chlorofluorocarbons—freons and other gases now thought to be depleting the ozone layer—produce a greater Southern Hemisphere signal. The concentrations of these manmade compounds in the atmosphere have been increasing rapidly in the past several decades. They dissolve in the surface layer of the ocean and mix downward slowly. John L. Bullister of the Woods Hole Oceanographic Institution has recently discovered a form of freon called freon 11 in a prime source of the Antarctic bottom water: the Filchner Ice Shelf cataract. The Filchner Ice Shelf lies in the Weddell Sea directly south of the Atlantic Ocean and off the coast of Antarctica. Cold, dense water spills off the Filchner Ice Shelf into the deep Weddell Sea. The fact that Bullister measured high concentrations of freon 11 at a depth of 1,500 meters in the Antarctic is a clear indication that manmade compounds are beginning to enter the Antarctic bottom water.

MAJOR OCEAN CATARACTS are indicated on a map of the Atlantic. The Denmark Strait, which lies between Greenland and Iceland, produces what is probably the world's largest cataract, with a flow rate of about five million cubic meters per second. The Iceland-Faroes cataract supplies the eastern North Atlantic with cold, dense water. The Ceara Abyssal Plain cataract, flowing from south to north, supplies the North Atlantic with its coldest and deepest water—the Antarctic bottom water. Through Discovery Gap water flows at a rate of 210,000 cubic meters per second from the eastern equatorial Atlantic to the eastern North Atlantic. The Filchner Ice Shelf cataract produces some of the world's densest bottom water. The South Shetland Islands cataract may be important in maintaining krill breeding grounds. Unlike these cataracts, which are due to temperature differences, the Strait of Gibraltar cataract is driven by differences in salinity. The red line traces the part of the 1972–73 GEOSECS expedition that resulted in the slice shown on the next two pages.

The most spectacular manifestation of giant ocean cataracts is of course their enormous flow rates. The Denmark Strait cataract once again provides the prime illustration. In 1967 L. Val Worthington of Woods Hole attempted to measure its flow rate by deploying an array of 30 current meters at various depths in the sill region. The currents were so severe that 20 of the meters were never recovered. Those that were recovered had recorded currents of up to 1.4 meters per second, which is sizable compared with the rates—from .1 to .5 meter per second—at which surface currents usually flow. It was on the basis of this peak current that the volume flux of the Denmark Strait cataract was estimated at five million cubic meters per second, the figure I gave above.

Unfortunately since Worthington's initial studies few flux measurements have been made. In 1973 workers of the Bedford Institute of Oceanography in Dartmouth, Nova Scotia, arrived at an estimate of 2.5 million cubic meters per second. The threat posed to moored current meters by the deep-sea trawlers in the Denmark Strait may preclude any future surveys.

In 1978 Worthington and I began measurements of the flow rate into the Ceara cataract, which delivers Antarctic bottom water to the North Atlantic Basin. Current meters deployed on the sill at four degrees north latitude yielded an estimated flow rate of from one to two million cubic meters per second—or some five to 10 times the flow rate of the Amazon. By this time buoy technology had advanced, so that we were able to deploy the moorings for a full year without any losses. This enabled us to measure the flux at tenth-of-a-degree intervals between 1.0 and 1.9 degrees C.

At about the same time Nelson G. Hogg and William J. Schmitz, Jr., of Woods Hole, in conjunction with Wilford D. Gardner and Pierre E. Biscay of the Lamont-Doherty Geological Observatory, measured the flux into the Rio Grande cataract. After two years of observations they estimated the flow rate into this cataract to be four million cubic meters per second, perhaps as large as the Denmark Strait cataract.

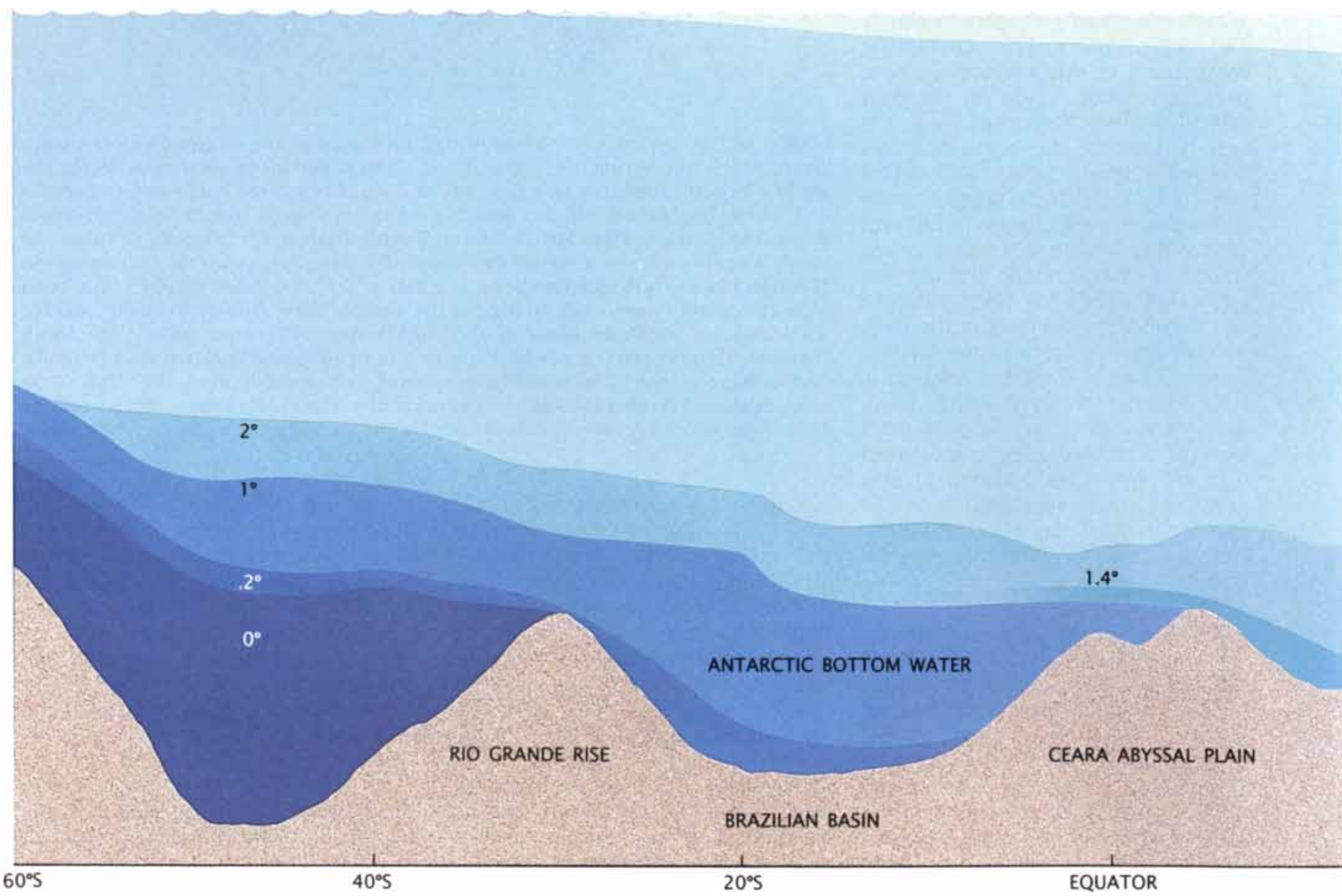
These enormous fluxes are generally driven by temperature differences between two basins, but not always. For example, owing to evaporation, water in the Mediterranean is much saltier—and therefore much denser—than even the deepest water in the Atlantic, even though it is warmer. Water flowing out of the Mediterranean through the Strait of Gibraltar therefore tends to sink into the Atlantic as a cataract.

During its descent the salty Mediterranean water mixes with Atlantic water. This mixing reduces the Mediterranean water's density until at a depth of only 1,000 meters it equals that of the surrounding Atlantic water. At this depth the water ceases its descent and spreads out in a plume over a sizable part of the northeastern Atlantic, where it is responsible for a pronounced maximum in ocean salinity.

The turbulent mixing of the Gibraltar cataract with Atlantic waters explains the depth of the plume but is itself a puzzle. No one yet knows why the Gibraltar cataract mixes more strongly with its surroundings than the Denmark Strait cataract does. Four potential explanations are (1) the steepness of the continental slope off Spain, (2) rougher geologic features there than in the Denmark Strait, (3) the fact that the Mediterranean cataract is driven by a difference in salinity whereas the Denmark Strait is a thermal cataract, and (4) the large surges in the flow through the Strait of Gibraltar as a result of tides and storms. To determine which if any of these explanations is correct will require fundamental studies in geophysical fluid dynamics in conjunction with much more observation.

Even though the factors determining the degree of mixing are highly uncertain, investigators agree that mixing is extremely important in determining the heat budgets of cataracts and their surroundings. The Ceara cataract protrudes into the North Atlantic as a tongue of cold water. Because the tongue has been observed to be stationary over many years, the northward-flowing water must be warming and hence rising through the isotherms. That is possible only if warmer water from above is mixing into the cataract water and heating it. Fluid dynamicists call this process turbulent-eddy mixing.

Indeed, the sharp boundaries of cataracts make them a good place to study turbulent-eddy mixing in the ocean. The degree of mixing and so the thermal-energy budgets (the degree to which heat can be transferred in and out of a cataract) are relatively easy to obtain. Studies of cataracts' thermal-energy budgets by Hogg, Schmitz, Gardner and Biscay, by Worthington and me and by Peter Saunders of the Institute of Oceanographic Sciences in England all in-



NORTH-SOUTH SLICE of the Atlantic from Greenland to Antarctica reveals several cataracts. The Rio Grande Rise at about 30 degrees south latitude blocks the northward flow of Antarctic bottom water with a temperature below zero degrees Celsius. Water at .2 degree, however, spills over the rise into the Brazilian Basin, forming the Rio Grande cataract. The Ceara Abyssal Plain near the Equator prevents water colder than one degree from flowing farther north. Water at 1.4 degrees spills

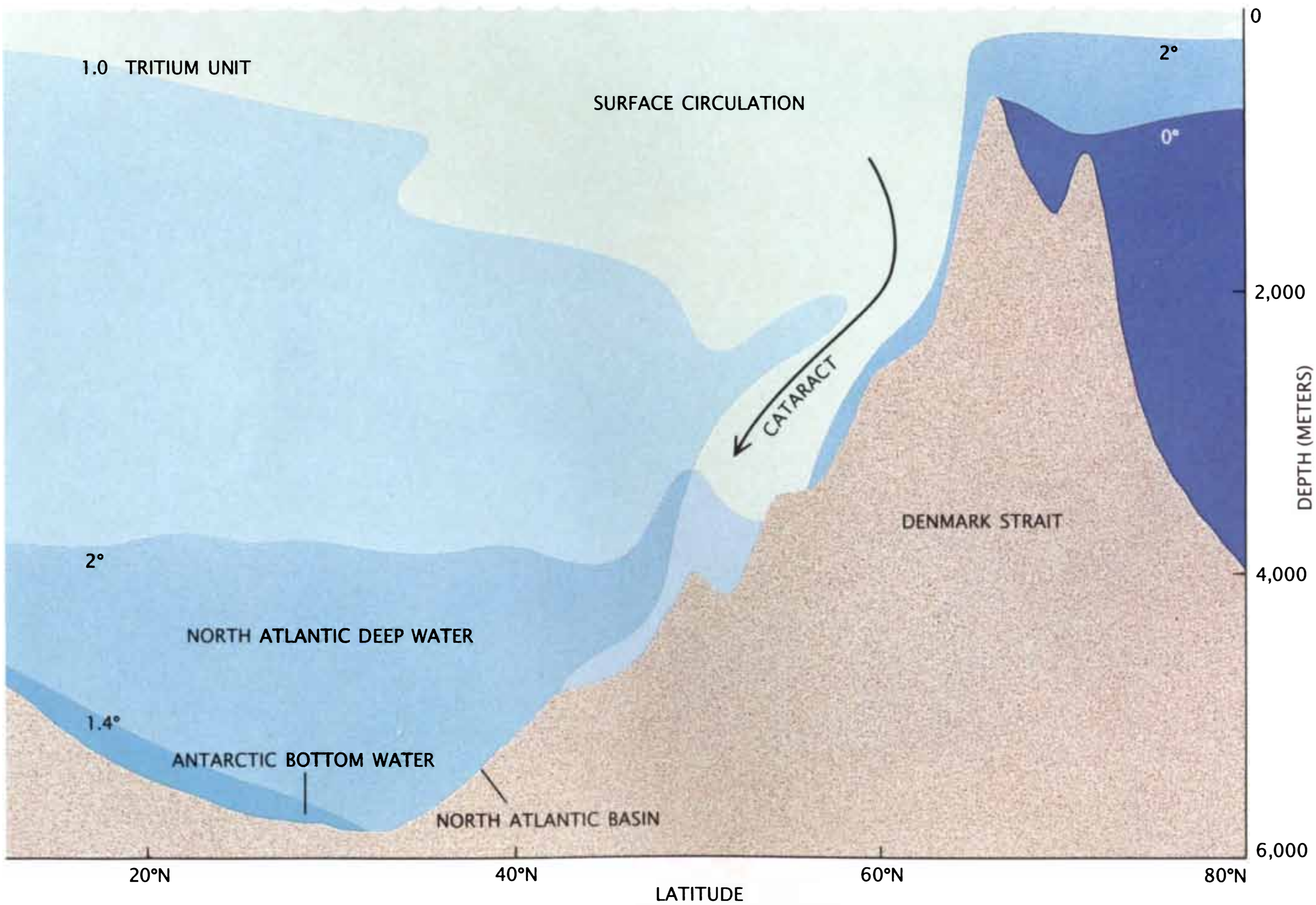
dicate that turbulent mixing is approximately 1,000 times as efficient in transferring heat as ordinary molecular conduction. Eventually we hope to determine thermal-energy budgets for other cataracts and ultimately for the ocean as a whole.

Since to some extent a cataract can be viewed as a system isolated from the rest of the ocean, it lends itself to laboratory study. In 1959 Thomas H. Ellison and J. Stewart Turner, then at the University of Manchester, constructed a simple experiment: they released salty water at the top of an incline that was itself at the bottom of a tank filled with fresh water. The denser salt water sank to the bottom of the tank, and the change in salinity as a function of the incline's slope and the flow rate was measured.

The slope and flow rate determine the Froude number, which essentially measures the ratio of inertial forces to buoyancy forces (the ratio of the actual velocity of an object to the velocity it would acquire if it were subjected only to gravitational acceleration). Ellison and Turner found that when the Froude number of such a laboratory cataract was much greater than 1, the flow became turbulent; the resultant mixing in turn drove the Froude number toward 1. This suggested, then, that the flow of cataracts tends to balance inertial and buoyancy forces.

The Ellison and Turner experiment was only a crude model of an ocean cataract and ignores a number of complicating factors that must be included for a more realistic description. Principal among these are the Coriolis force and friction. The Coriolis force is due to the rotation of the earth; it tends to deflect a moving object at a right angle to the object's trajectory. If the object is moving along a north-south path, it will be deflected to its right in the Northern Hemisphere and to the left in the Southern Hemisphere. In the Denmark Strait cataract, Coriolis forces cause the water to rise about a kilometer higher on the right-hand wall of the channel than on the left-hand wall. Also because of Coriolis forces, the salty Mediterranean current bends to the right after leaving the Strait of Gibraltar; the cataract it forms actually flows parallel to the Spanish coast.

An attempt to take certain of the complicating factors into account in analyzing the dynamics of ocean cataracts was made by Peter C. Smith of the Massachusetts Institute of Technology in 1973. Smith constructed a laboratory model that produced the Coriolis force by means of a rotating turntable, and that also included bottom friction (friction between the water and the channel bottom). Then, in a theoretical model, he added friction between the cataract and overlying layers of water, which was taken into account in the form of entrainment, or the mixing of different fluid layers. From the theoretical



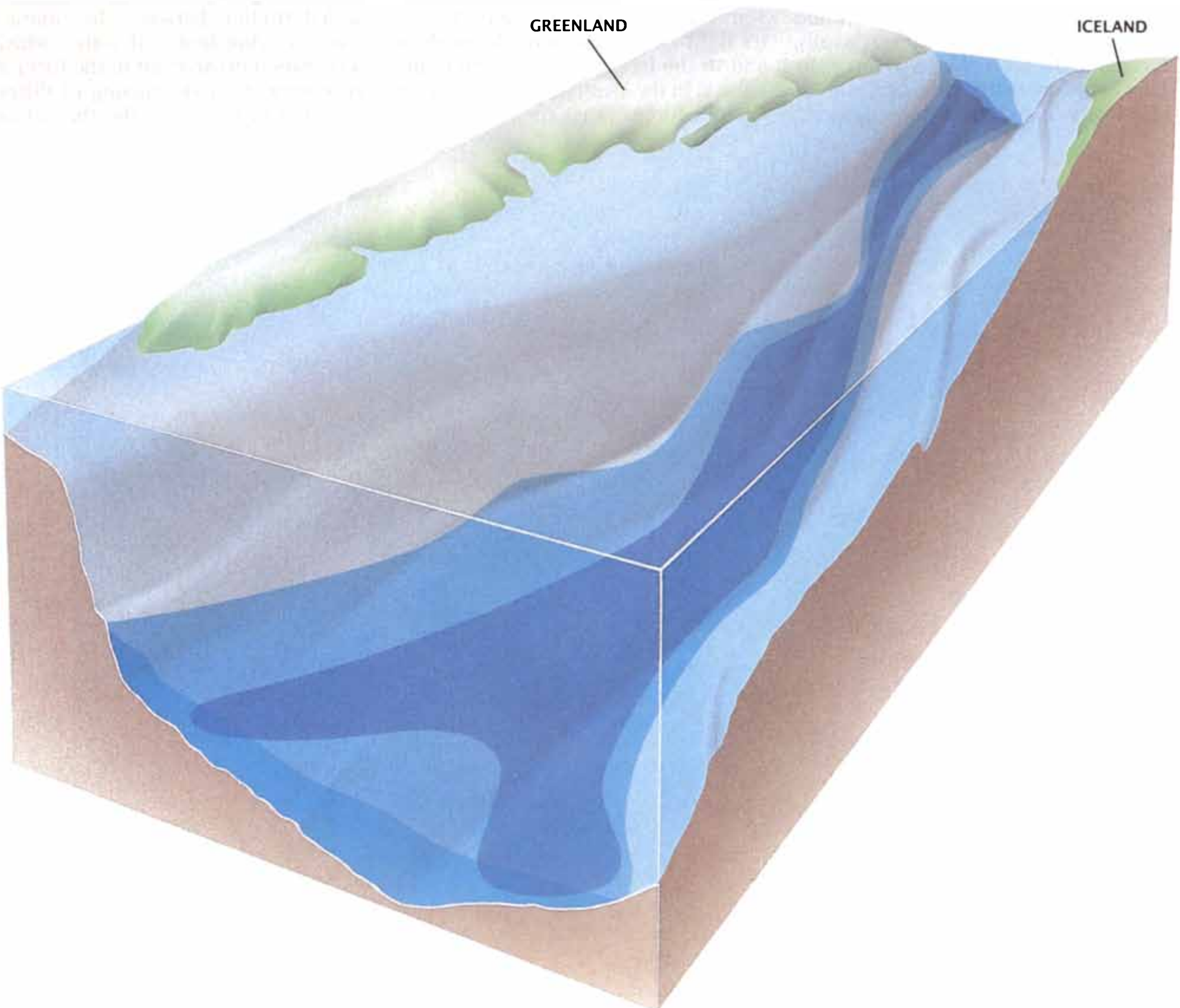
over the rise into the North Atlantic Basin as the Ceara cataract, supplying the North Atlantic with Antarctic bottom water. The Denmark Strait cataract, flowing southward, carries water at two degrees from the Norwegian Sea down to roughly 3,500 meters; it forms the North Atlantic deep water, which lies just above the Antarctic bottom water. In addition to temperature, tritium, a radioactive isotope of hydrogen produced in atmospheric nuclear-weapons tests, also traces this cataract.

model he concluded that the mixing rates found in the Denmark Strait are the result of bottom friction rather than drag between neighboring water layers. He also concluded that bottom friction dominated near the top of the cataract, whereas entrainment became important farther downstream.

In the past few years James F. Price and Martha T. O'Neil of Woods Hole have refined Smith's model; their formula for entrainment is based on high-resolution studies of mixing in the upper ocean. Working with Thomas B. Sanford of the University of Washington and Rolf G. Lueck of Johns Hopkins University, they have made measurements of the velocity, temperature and salinity within the Gibraltar Strait cataract. Comparison of the observations with the model indicates that it gives a realistic picture of the density change along an ocean cataract. Their findings also suggest (lending confirmation to Smith's results) that drag is caused mainly by bottom friction unless a certain critical velocity is exceeded, at which point the water mixes strongly with its surroundings. In contrast to Smith's result, however, the workers find that entrainment is important near the top of the cataract, whereas bottom friction is important farther downstream.

There is still a question, however, about whether the results of theoretical models and laboratory experiments—which are limited by the size of the turntables—can be scaled up to the real ocean, where turbulence is likely to be much greater and where the topography of the bottom introduces a largely unknown amount of drag and mixing. Indeed, the degree of drag and mixing at the ocean bottom has only recently begun to be measured. Consequently it is still not clear that the amount of friction assumed by investigators is appropriate to the real ocean; new data will help to decide whether the assumptions that are built into the models are reasonable or need to be discarded.

In spite of the scarcity of good data, investigators continue to increase the sophistication of the models. Recent laboratory experiments have elucidat-

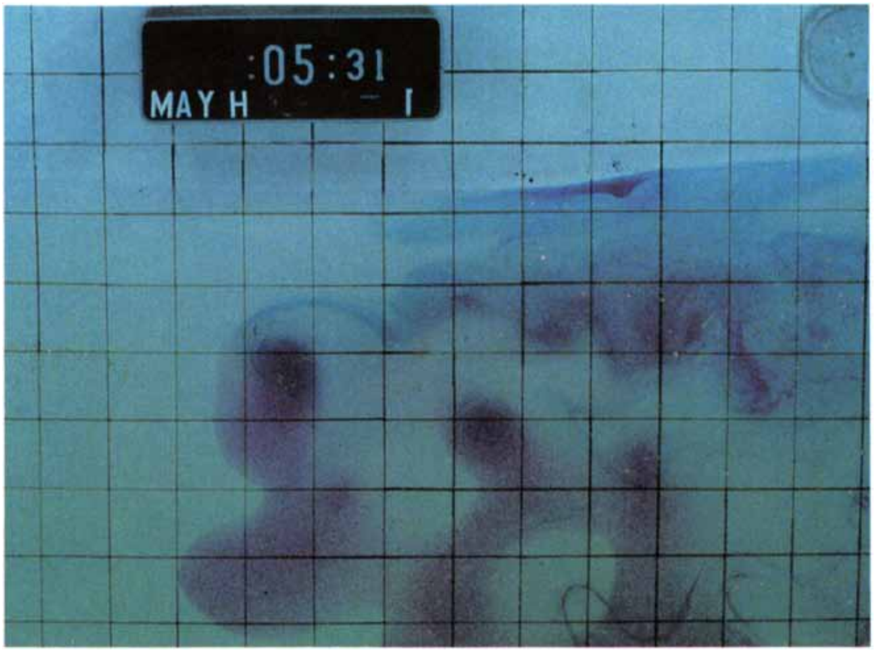


DENMARK STRAIT CATARACT is shown in perspective. The sill of this cataract is the Iceland-Greenland Rise, which lies 650 meters below sea level. Water from the Norwegian Sea flows over the rise at a rate that has been estimated to be 25 times that of the Amazon. The Coriolis force causes the water to rise about a kilometer higher on the right-hand wall (looking downstream) than on the left-hand wall. The dark blue contour is one-degree water, which fans out around the tip of Greenland. Above it lies 1.8-degree water, which protrudes into the North Atlantic as a tongue extending as far south as Newfoundland.

ed the effects of the Coriolis force in more detail. For instance, experiments done on a turntable by Ross W. Griffiths of the Australian National University and independently by Melvin E. Stern of Florida State University, Glenn R. Flierl and Barry A. Klinger of M.I.T. and me show that the Coriolis force results in the generation of isolated eddies, or small ocean whirlpools, directly above a cataract near the bottom of the ocean.

Two contrasting mechanisms of eddy generation have been proposed. Griffiths suggests the cataract produces an "inertial" wave in the ocean, that is, a vorticity wave that rotates first in one direction and then in another, like the water in a washing machine. The waves radiate upward and, according to Griffiths, they break and produce intense cyclones above the cataract. Such a process has indeed been observed in experiments in which a stirring grid produces inertial waves in a spinning fluid.

Stern, Flierl, Klinger and I propose a different mechanism. We think cyclonic circulation is a direct product of vertical mixing from small-scale turbulence in a rotating fluid when dense water lies under less dense water. According to this model, the turbulence mixes the less dense water above with the denser water below. The density of the former is effectively increased and it sinks, sucking water directly above it downward. The Coriolis force reorients the flow into a horizontal plane, turning it into a vortex, which lies between the cataract and the water's surface. Such eddies could be counterparts of the larger ocean eddies that can be several hundred kilometers across and persist for several years [see "Rings of the Gulf Stream," by Peter H. Wiebe; SCIENTIFIC AMERICAN, March, 1982], but to date the smaller vortexes have not been observed outside the laboratory.



LABORATORY EDDIES are created by introducing a blob of dense salt water, here dyed red, into a rotating tank of fresh water, some of which is dyed blue. The dense water descends, mixing turbulently with the surrounding water and drawing down the less dense water above. The Coriolis force makes the downward motion cyclonic, generating eddies. Three eddies can be seen here, each consisting of a lens of red fluid overlain by blue surface water. The same process, triggered by the descent of dense water in cataracts, may generate ocean eddies that can be kilometers across.

In addition to their effects on the climate and salinity of the deep ocean, cataracts might have important effects on marine biology. One such effect can be seen in the South Shetland Islands cataract. After leaving the Weddell Sea, which is south of the Atlantic Ocean, the South Shetland Islands cataract eventually descends westward toward the the Scotia Sea, south of the Pacific Ocean. Worth D. Nowlin, Jr., of Texas A&M University and Walter Zenk of the Institute of Marine Sciences at the University of Kiel in West Germany have suggested that the life cycle of krill in the Scotia Sea is shaped by the deep cataract.

Biologists know that the adult krill gather to spawn near the South Shetland Islands, forming extensive whale-feeding grounds. These spawning grounds lie near the immense Antarctic circumpolar current, a surface flow that sweeps eastward. Krill eggs are laid at depths of approximately 50 meters but sink to about 1,000 meters before hatching, and it has been a puzzle how the larvae, which cannot swim, can return to the spawning grounds as adults.

Curiously, it has been observed that the nursery grounds, where the larvae are found, are located hundreds of kilometers to the west, even though the Antarctic circumpolar current flows to the east. Nowlin and Zenk suggest that bottom cataracts are carrying the krill westward. If this scenario is correct, then the larvae rise after hatching and are carried eastward again to the spawning grounds by the Antarctic current. Thus ocean cataracts may provide the mechanism to close the circuit between the spawning area, the nursery and the feeding grounds.

Although much has been learned over the past 20 years about the extent of cataracts and their flow rates, temperature profiles and chemical contents, much more study needs to be done in order to determine the rates of mixing with surrounding water and the amount of drag at the ocean bottom. The result will be a clearer picture of the largest water cascades on the planet, which through their influence on the salinity, temperature and biology of the ocean have an effect on the climate and ecology of the entire earth.

FURTHER READING

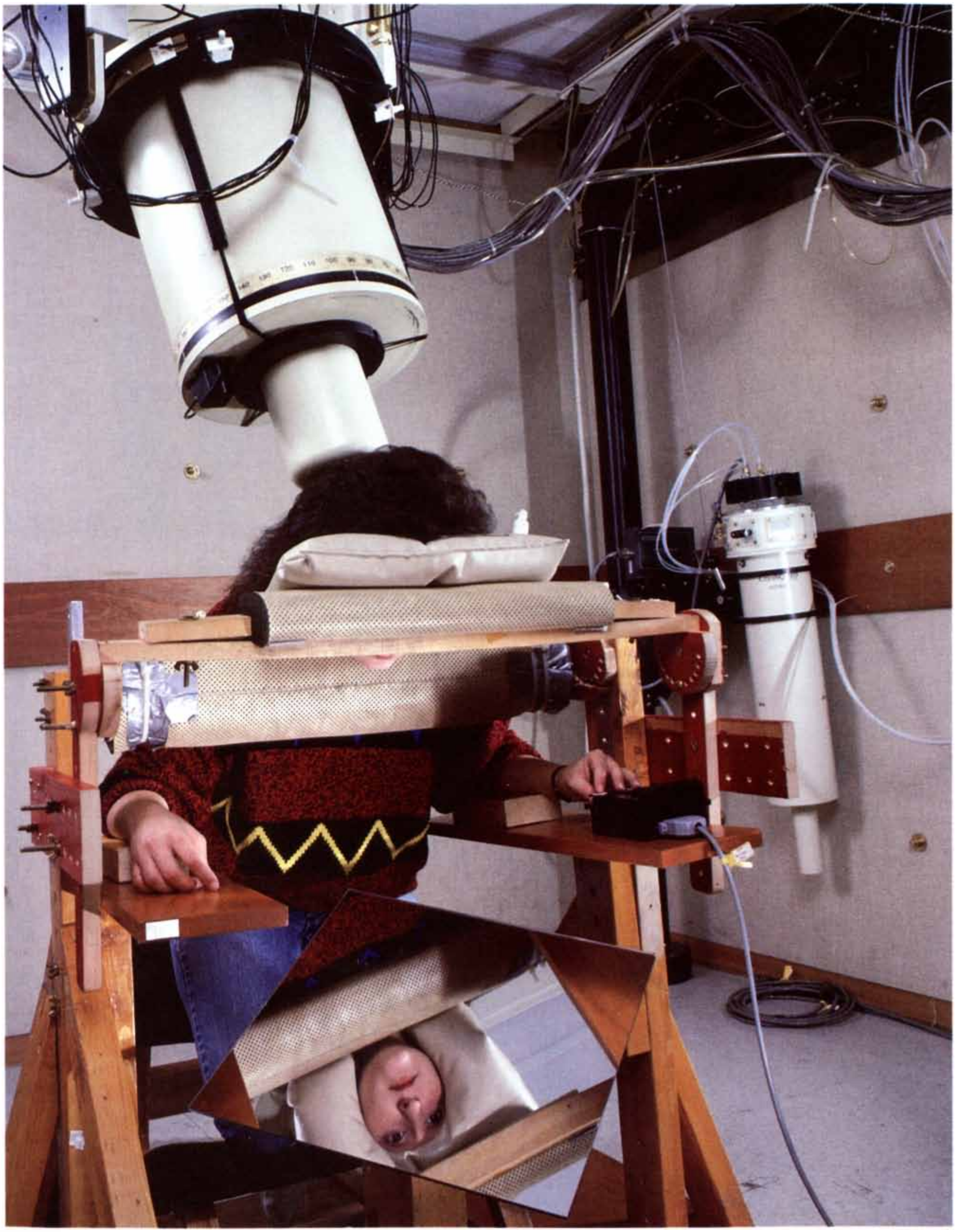
AN ATTEMPT TO MEASURE THE VOLUME TRANSPORT OF NORWEGIAN SEA OVERFLOW WATER THROUGH THE DENMARK STRAIT. L. V. Worthington in *Deep-Sea Research,* Vol. 16, Supplement, pages 421–432; August 1, 1969.

A STREAMTUBE MODEL FOR BOTTOM BOUNDARY CURRENTS IN THE OCEAN. Peter C. Smith in *Deep-Sea Research,* Vol. 22, No. 12, pages 853–873; December, 1975.

THE FLUX AND MIXING RATES OF ANTARCTIC BOTTOM WATER WITHIN THE NORTH ATLANTIC. J. A. Whitehead, Jr., and L. V. Worthington in *Journal of Geophysical Research,* Vol. 87, No. C10, pages 7903–7924; September 20, 1982.

GEOSECS ATLANTIC, PACIFIC, AND INDIAN OCEAN EXPEDITIONS, VOL. 7: SHORE-BASED DATA AND GRAPHICS. GEOSECS Executive Committee, H. Gote Ostlund, Harmon Craig, Wallace S. Broeker and Derek W. Spencer. National Science Foundation, U.S. Government Printing Office, 1987.


DOHC
16VALVE



SUPERCONDUCTING quantum interference devices (SQUIDS) inside the white cylinder placed against the back of a woman's head detect magnetic fields induced by her brain's electrical activity as she responds to visual cues viewed in a mirror. In contrast to electric fields, the magnetic fields are not distorted by the skull. The technique, which is called magnetoencephalography, may prove to be a significant new application of low-temperature, and perhaps also high-temperature, superconductors. The photograph was made in the laboratory of Samuel J. Williamson and Lloyd Kaufman at New York University.

The New Superconductors: Prospects for Applications

Further feats of materials science are needed to realize commercial hopes for high-temperature superconductors. A long-term commitment to research will make success more likely

by Alan M. Wolsky, Robert F. Giese and Edward J. Daniels

In the 80 years since Heike Kamerlingh Onnes discovered superconductivity, much has been learned about superconductors—and yet Onnes' initial thoughts about practical applications remain relevant. In 1911 he was astonished to find that mercury cooled by liquid helium to four degrees Kelvin lost all electrical resistance; two years later, after observing superconductivity in tin and lead at 4 K, he wrote: "Tin and lead being easily workable materials, we can now contemplate all kinds of electrical experiments with apparatus without resistance.... The extraordinary character of this [superconducting] state can well be elucidated by its bearing on the problem of producing intense magnetic fields with aid of coils without iron cores." Iron cores have two drawbacks: they are heavy and they have only limited ability to magnify the magnetic field of the current in the surrounding coil.

The hope for superconducting magnets without iron cores was frustrated, however, because superconducting tin and lead could not transport large enough currents. Fifty years elapsed before investigators discovered that niobium-titanium alloy and niobium-3-tin can carry the needed currents. But like tin and lead, they also had to be cooled to four degrees K (four degrees above absolute zero) with expensive helium, and that greatly limited their applications. Then, two years ago, several groups around the world, excited by the discovery by K. Alex Müller and J. Georg Bednorz of the IBM Zurich Research Laboratory of a superconducting ceramic oxide, developed an yttrium-barium-copper oxide that superconducted at 90 K. Since then other investigators have found two separate families of copper oxides, one incorporating bismuth and the other thallium, that superconduct at between 110 and 120 K.

These high-temperature superconductors could be cooled to 77 degrees K with liquid nitrogen, which is cheap and abundant. This immediately suggested that certain applications of superconductivity long considered not to be economic or practical might be feasible. Yet many of the envisioned applications—generators and motors, energy storage, magnetically levitating trains—raise the same issues to which Onnes referred. It is not yet known whether the new materials can be made "easily workable"—strong and flexible enough to fashion into wire and other useful forms. Nor is it known whether they can be made to carry large currents and operate in intense magnetic fields.

Whether the new discoveries will prove fruitful will depend on the progress made toward achieving design requirements for known applications and on identifying new applications as yet unforeseen. Indeed, such new applications may well have the greater impact. No one foresaw today's most important commercial use of superconductivity, magnetic-resonance imaging for medical diagnosis, in the 1960's, when niobium-3-tin and niobium-titanium were found to remain superconducting while carrying high currents in the presence of sizable magnetic fields. Leaving aside the unforeseen, an informed view of the economic and technical advantages of the new superconductors can help guide attempts to achieve the applications now being envisioned.

ALAN M. WOLSKY, ROBERT F. GIESE and EDWARD J. DANIELS are in the technology evaluations group of the energy and environmental systems division at the Argonne National Laboratory. Wolsky is senior energy scientist and director of the group. He got his Ph.D. in physics from the University of Pennsylvania in 1969 and was visiting assistant professor at Temple University before joining Argonne. Giese, an energy-systems analyst, went to Argonne soon after he got his Ph.D. in physics from Stanford University in 1974. Daniels earned an M.B.A. at the University of Chicago in 1981. He joined Argonne in 1985 as an energy-systems analyst. The authors thank the Office of the Deputy Assistant Secretary for Renewable Energy of the U.S. Department of Energy for its support.

The Cold Facts

Many elements can be either normal conductors or superconductors, just as water can be liquid or solid. And just as there is a particular temperature above which water is liquid and below which it is solid, so there is a particular "transition temperature" above which a substance is normal and below which it becomes superconducting. Pressure affects the transition temperature, just as high pressure raises the freezing point of water. In a superconductor, electric currents and magnetic fields also alter the transition temperature—effects crucial to the feasibility of applications.

Normal materials have electrical resistance, a kind of "friction" that causes electrons in a current to dissipate their energy in the form of heat. A superconductor has zero resistance and so can sustain a direct current without losing power or generating heat. (Alternating currents lose a small amount of power that is proportional to the a.c. frequency and is dependent on the size and shape of the wire, the condition of its surface and the magnetic field strength.)

The same mechanism that causes a superconductor to lose electrical resistance also causes it to become responsive to magnetic fields. If a superconductor is placed in a modest magnetic field and the temperature is lowered below the transition point, it will abruptly expel the field, acting as though it contains tiny bar magnets that diminish the external field. This is the famous Meissner effect, which causes a lump of superconductor to be levitated above a magnet. The same property means that superconductors could shield against magnetic forces, just as electrical conductors shield against electrostatic forces.

Why is it better to have superconductivity at 77 degrees K than at 4 K? It is less expensive. To keep a superconductor cold one must continually replace the liquid helium (4 K) or nitrogen (77 K) vaporized by heat leaking in from outside. In one hour one watt of heat will evaporate 1.4 liters of liquid helium or .016 liter of liquid nitrogen. To replace this loss would cost $50,000 every year for liquid helium (priced today at $4 per liter), compared with $35 for liquid nitrogen (at 25 cents per liter). For large-scale applications in which refrigeration is a minor part of the cost, switching to liquid nitrogen would not make much of an economic dent. The switch to liquid nitrogen would have a bigger impact on the cost of medium-size and small applications, which are comparatively expensive to cool with helium because of their large surface-to-volume ratio. Here the savings could be a significant fraction of total cost. This is important because small devices are usually cheaper than large ones and so are more likely to be tried first.

What is more, with liquid-helium-cooled systems, engineers often design complex thermal insulation and helium-recycling systems to avoid wasting the expensive coolant. Such equipment is itself costly, and the added complexity diminishes the reliability of the entire system. With liquid nitrogen, thermal insulation could be made less effective and thus less expensive, and the vapor-recovery system could be abandoned, yielding greater simplicity and reliability.

Supermagnets

Magnets are essential to every electric generator and motor. In industry magnets lift and separate iron and steel products and draw impurities out of clays and other materials. The latest medical diagnostic tool, magnetic-resonance imaging, uses a superconducting magnet to align hydrogen nuclei in the body; radio pulses jostle the nuclei, which, like toy dolls that stand up when pushed over, wobble back into alignment. As they wobble, the nuclei emit weak radio signals of their own that can be detected.

Most electromagnets are made by winding insulated copper wire around an iron-alloy core. Current flows in the coil, inducing a magnetic field that points along the coil's axis. By itself the coil can generate only a modest field, because copper wire carries only about 400 amperes per square centimeter. (If more current is forced through the wire, the cost of removing the heat generated by the wire's resistance becomes prohibitive.) The iron-alloy core greatly amplifies the coil's magnetic field. Electrons in the core align their magnetic moments with the coil's field, enhancing its strength hundreds or thousands of times, thus making it economic to deliver fields of up to two teslas—40,000 times the earth's field. Stronger fields are impossible because the core "saturates," or runs out of alignable electrons.

Iron-core electromagnets have another disadvantage: they are heavy. The iron accounts for much of the weight of electric motors and generators. If one were to reduce equipment weight (and thereby increase payload) in aircraft, the reduction might be worth up to $500 per pound saved over the aircraft's lifetime. The savings are even greater—about $3,000 per pound—for Titan 4 rockets launching payloads into low earth orbit.

Lighter, more powerful magnets are possible with superconductors. In the simplest design one can omit the iron core and have the coil carry a proportionally larger current. If an iron core magnifies by 1,000 times the field generated by a copper coil, then to match the performance of the conventional magnet, the superconductor must carry 1,000 times as much current as the copper wire, or about 400,000 amperes per square centimeter. Niobium-titanium and niobium-3-tin achieve these current densities, but only at four degrees K. Bulk samples (but not as yet wire) of yttrium-barium-copper oxide have been made at the AT&T Bell Laboratories that carry 4,000 amperes per square centimeter at 77 K in a one-tesla magnetic field. Without the field the samples carried 17,000 amperes per square centimeter.

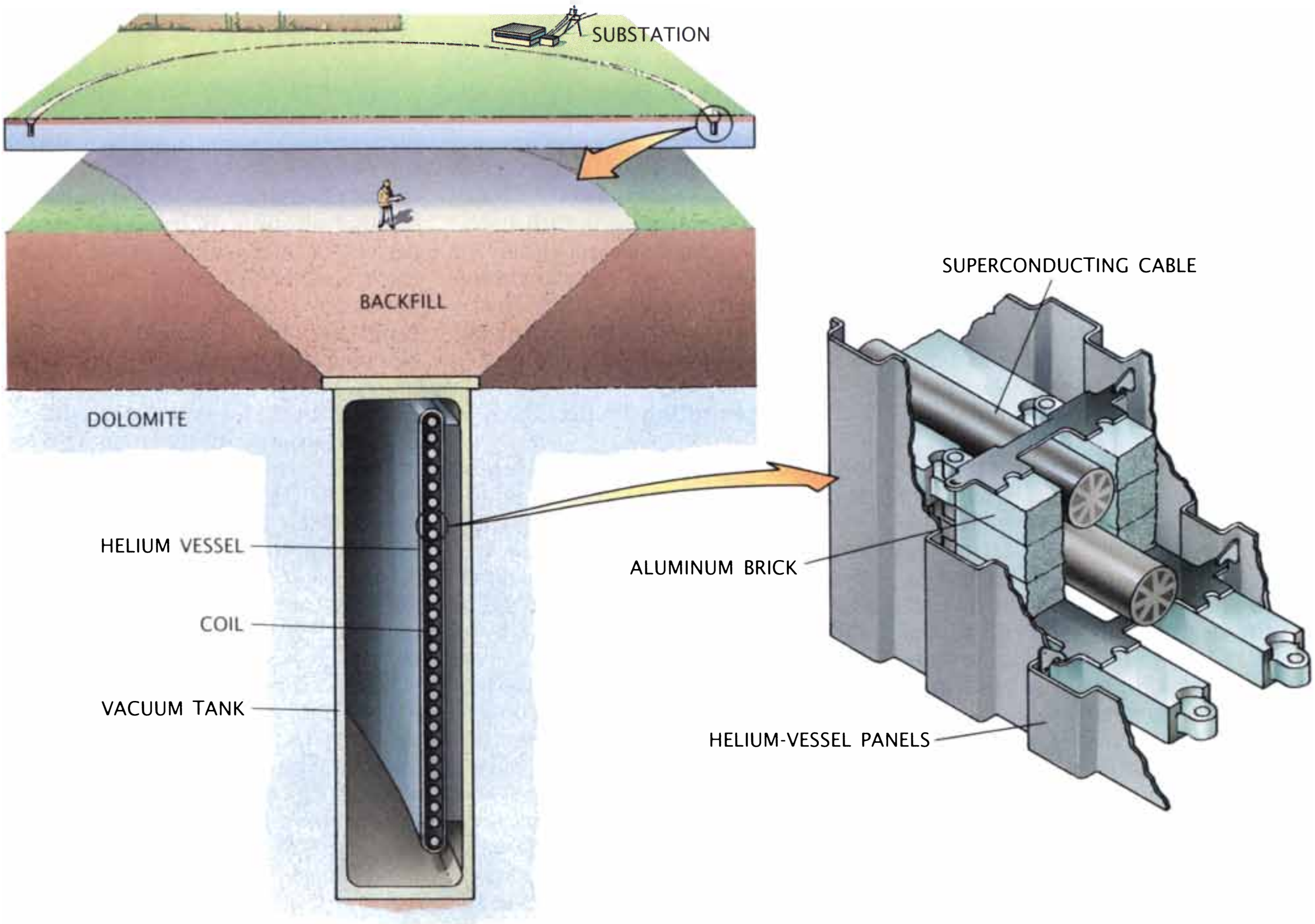
Generators and Power Lines

In electric generators mechanical power provided by a spinning turbine rotates a magnet, which induces an electric current. Superconducting magnets could increase the efficiency of such generators [see "Superconductors in Electric-Power Technology," by T. H. Geballe and J. K. Hulm; SCIENTIFIC AMERICAN, November, 1980]. Small-scale generators using low-temperature superconductors have been built in the U.S. and Japan. Large superconducting generators are projected to be about 99.5 percent efficient; conventional generators, however, are already 98.6 percent efficient. The annual saving in fuel cost would be about 1 percent—significant compared with the annualized cost of the generator (an investment of several million dollars) but dwarfed by the capital investment in the rest of the electric power plant, which would be about 100 times as great.

Design studies have, however, revealed an entirely different benefit of superconducting generators, one that is difficult to quantify in monetary terms but is valuable nonetheless. Electric utilities face a serious problem when a short circuit occurs unexpectedly, for example when a lightning strike grounds a transmission line. Generators at nearby sites thereupon rotate faster, destroying their harmony with the rest of the grid. The utility must isolate the short circuit quickly and bring the generators back into phase before they spin out of control. Because it will be easier to regain control of air-core magnets than of iron-core magnets, the utility would have more time to remedy the problem.

Technologies for building superconducting turbine-generators have existed for a decade. Why, then, are none in operation? The answer is simple. If liquid helium is the coolant, refrigeration economies of scale make only large generators economic. But U.S. demand for new large generators has vanished, and so there has been no incentive to explore an unproved technology. The market for small generators (less than 100 megawatts) still remains, and here machines made of the new superconductors could compete with conventional ones. There are two reasons: the refrigeration systems would be less expensive and the system as a whole would be more reliable.

Similar considerations apply to the use of superconductors to transmit and distribute electric power. Present-day systems made of aluminum or copper dissipate between 5 and 8 percent of the power generated by U.S. utilities before it reaches the customer. On the other hand, copper and aluminum can be drawn into wire that is strong and flexible enough to be strung between towers. Air cools the



SUPERCONDUCTING MAGNETIC ENERGY STORAGE (SMES) system consisting of giant rings of superconducting cable could store a direct electric current indefinitely. A 5,000-megawatt-hour, 1,000-megawatt system would have to be buried to contain the magnetic force generated in the coil. A coil 24 meters high and 1,568 meters in diameter would impart a force of 15,000 pounds per square inch to the cavern wall. The cables are encased in very pure aluminum brick. In an emergency "quench," or breakdown, of superconductivity the brick would absorb the unleashed electricity and heat.

wire and insulates it electrically at no extra cost. A "room temperature" superconductor that could be strung overhead would be extremely valuable, but no such material yet exists. Superconducting transmission lines cooled with liquid helium or nitrogen would require heavy, rigid containment systems to support the necessary thermal and electrical insulation.

Such systems would have to be at ground level or, more likely, buried, and so they would cost at least three times as much to build as conventional overhead lines. These superconducting transmission lines will never compete with overhead lines. Underground superconducting transmission lines might, however, compete with underground conventional lines. Utilities now bury transmission lines for aesthetic reasons, but as people become concerned about the health effects of overhead lines, demand for buried systems may increase. The Brookhaven National Laboratory has constructed a 1,000-megawatt ground-level a.c. transmission line 115 meters long of niobium-3-tin cable cooled with helium, showing that the concept is feasible.

The economies of scale of refrigeration make liquid-helium-cooled transmission lines economic only if they carry very large amounts of power. In 1977, when the Philadelphia Electric Company studied alternative ways to build a 10,000-megawatt line into Philadelphia, superconducting systems proved to be roughly competitive. No demand for such large-capacity lines has materialized, however. Even if the demand existed, no utility would risk so much on a design that was unproved. High-temperature superconductors may make it possible to build smaller, more reliable lines.

Superconductors may find a different application in a device to protect power grids against unexpected—and unwelcome—surges. The device consists of a segment of transmission line wound around an iron-alloy core. The core is also wound with superconducting wire carrying a small direct current, which keeps the core fully saturated. If there is a sudden surge in the transmission line, the device will prevent the surge from propagating to undamaged parts of the grid. High-temperature superconductors might make it economic to have such surge protectors throughout a system.

Storing Electrical Energy

The demand for electric power is often out of synchrony with its availability. Coal-fired and nuclear power plants, which operate most efficiently at a level output, must serve demands that vary over the course of the day. An energy-storage system would enable a utility to store excess energy during times of overproduction and then tap it when demand exceeds generating capacity. A system made of superconductors would be able to hold a massive amount of direct current with almost no energy loss.

A full-scale superconducting magnetic energy storage (SMES) unit has been designed to store about 5,000 megawatt-hours of electricity and to charge and discharge at the rate of 1,000 megawatts. An SMES unit would be from 90 to 95 percent efficient (the loss comes from switching between a.c. and d.c. and from energy consumed for refrigeration) and would have a switching speed of a fraction of a second; no other system matches it. Such a system, however, is estimated to cost $1 billion and would be a major engineering challenge.

SMES designs employing niobium-titanium cable and liquid helium call for a ring-shaped underground system with a radius of about 750 meters. To prevent a catastrophic release of energy if the superconductivity vanishes, the designs call for $100-million worth of very pure aluminum brick to serve as a sink for electricity and heat. A liquid-nitrogen-cooled SMES system would reduce the cost of refrigeration, lowering the capital cost by about 3 percent. (If the superconductor itself is inexpensive, the cost might come down another 5 percent.)

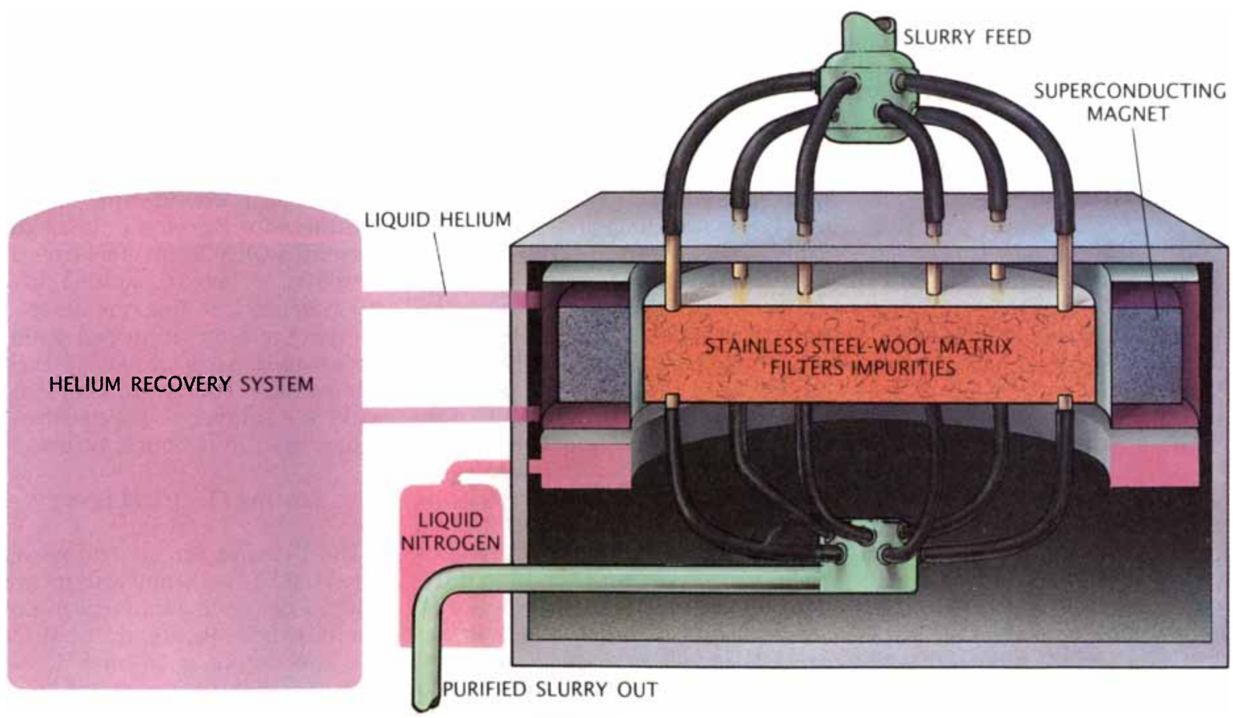
On a smaller and less expensive scale, SMES could be put to a different use: to smooth the transmission of power generated by highly erratic sources, such as an array of windmills. SMES units of this type must switch a large amount of power but do not have to store much energy, and so they do not require expensive containment systems. The Los Alamos National Laboratory operated a small experimental SMES unit cooled with liquid helium as a transmission buffer for the Bonneville Power Administration from 1983 to 1984. A failure in the refrigeration forced its retirement, but otherwise the SMES performed well.

Levitating Trains

One of the more widely discussed superconductor applications is the maglev (magnetically levitating) train, which would cruise at high speed, suspended above a guideway by magnetic forces. The idea, first proposed in the 1960's by James R. Powell, Jr., and Gordon T. Danby of Brookhaven, is already being developed. Germany's "Transrapid" employs conventional electromagnets and reaches speeds of up to 250 miles per hour, but the system is inherently unstable and must be controlled continuously by computers. The Japanese National Railways has been developing a stabler scheme with low-temperature superconducting magnets. As the train advances over aluminum coils in the guideway, the magnets induce opposing fields that levitate the train.

High-temperature superconductors would offer greater engineering reliability for maglev trains, but they would not reduce costs by much. To be sure, the refrigeration cost would go down, but that accounts for only a minuscule part of both the total capital investment and the operating cost. A 500-kilometer guideway would cost between $1.5 and $4.5 billion. The vehicles would cost only 10 percent as much, and the refrigeration system would account for only 1 percent.

Would maglev attract enough riders in the U.S. to justify such a huge investment? The vehicle's high speed might make it competitive with air travel for journeys of between 100 and 600 miles. If maglev were to link airports instead of city centers, it would not entail a change in travel habits, and yet it would help to relieve the air-traffic congestion and reduce the delays that are now routine at major airports. (Chicago's O'Hare International Airport, the nation's busiest, clocked more than 12 million passenger-hours of delay in 1986.) Investors must decide, however, whether it would be more profitable to relieve air-traffic congestion by maglev or by some alternative, such as in-



MAGNETIC SEPARATOR feeds a slurry of unprocessed material into a tank filled with a matrix of stainless-steel wool. As a magnet imparts an intense, uniform magnetic field, ferritic impurities are captured in the matrix. A system using superconducting magnets cooled with liquid helium consumes 95 percent less electricity than a conventional system. High-temperature superconductors cooled with liquid nitrogen would eliminate the helium-vapor recovery system, thereby increasing reliability and reducing the system's $2-million capital cost by roughly $200,000 and the operating cost by 10 percent.

creasing the number of passengers on a plane (by replacing 727's with 757's, for example).

Superconductors in Industry

Large-scale applications of superconductors in the electric-power industry and maglev have commanded the most visibility, but smaller-scale applications, particularly for various industrial machines and electronic devices, offer greater potential payoffs, particularly in the near term. Small-scale applications enjoy a market that is larger, has a faster turnover and is more responsive to innovation.

Powerful magnets are employed in industry to remove impurities from food and from raw materials such as alumina, calcite and sand. The first industrial low-temperature superconducting magnetic separator was installed in 1986 at the J. M. Huber Corporation's clay-processing facility. The system's two-tesla helium-cooled magnet requires about 60 kilowatts of electric power to operate the helium reliquefier, whereas a conventional water-cooled magnet would require 300 kilowatts to overcome electrical resistance and remove heat. Similar superconducting systems would cost about $2 million, compared with $1.6 million for a conventional one; savings in electricity should pay back the superconducting system's $400,000 added cost in two to three years. With high-temperature superconductors the helium reliquefier could be eliminated, saving 60 kilowatts and reducing capital cost by $200,000.

Superconducting magnets also offer a unique way to achieve extraordinarily low temperatures. Below their "Curie temperature" molecules in paramagnetic or ferromagnetic materials align themselves with magnetic fields, overriding the effects of random thermal motion. If such materials are put in a magnetic field, they heat up; when the field is removed, they cool down. This magnetocaloric effect can be exploited in a heat pump. Good performance, however, requires a field of about 10 teslas, which in practice can be produced only by a superconducting magnet. In the laboratory, magnetic refrigerators have achieved temperatures of 10^{-6} degrees K. They are now being developed for cryogenic applications in space and defense.

High-temperature superconductors could extend magnetic refrigeration to the industrial and commercial sector, particularly for food processing, where such devices promise to be more reliable, more compact, twice as energy-efficient and cheaper to operate over their lifetime than conventional refrigerators. What is more, as the chlorofluorocarbons that are now widely employed in conventional vapor-compression refrigerators are restricted because of their suspected role in the depletion of atmospheric ozone, magnetic refrigeration may provide a viable alternative.

LIFE-CYCLE DOLLAR SAVINGS (PERCENT)	COMPARED WITH LOW-TEMPERATURE SYSTEM	COMPARED WITH CONVENTIONAL SYSTEM
300-MEGAWATT GENERATOR	27	63
1,000-MEGAVOLT-AMPERE TRANSFORMER	36	60
10,000-MEGAVOLT-AMPERE, 230-KILOVOLT TRANSMISSION LINE	23	43
5,000-MEGAWATT SMES SYSTEM	5-8	VARIABLE
MOTORS	11	21
MAGNETIC SEPARATORS	15	20

LIFE-CYCLE DOLLAR SAVINGS (by percentage) of high-temperature-superconductor applications are compared with conventional and low-temperature-superconductor systems. High-temperature superconductors are assumed to match the electrical and mechanical performance of low-temperature niobium-titanium and niobium-3-tin superconductors and to cost the same. Practical, liquid-nitrogen-cooled superconductors would make the applications shown here economic at smaller scales. They would also make possible a diversity of new smaller-scale applications.

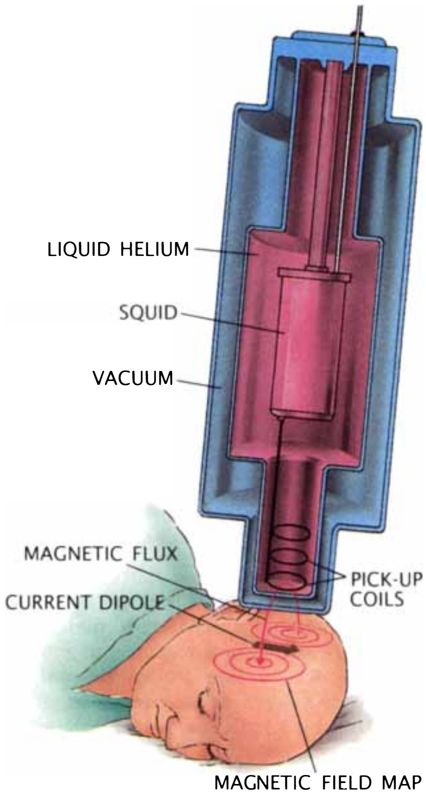
Superconducting Computers

Ever since the mid-1960's, when superconducting effects postulated by Brian D. Josephson of the University of Cambridge were realized in an electronic device that now bears his name, people have dreamed of building a superconducting computer. Josephson junctions, which consist of a thin layer of insulating material sandwiched between layers of a superconducting material, switch voltages very fast while consuming several orders of magnitude less energy than conventional devices. Their low power consumption promised more compact computers, because less internal space would have to be devoted to cooling. Compactness would enable faster computers to be built, because communication between different components is limited by the speed of light [see "The Superconducting Computer," by Juri Matisoo; SCIENTIFIC AMERICAN, May, 1980].

In a quest to build a superconducting computer, IBM spent nearly two decades developing Josephson-junction logic circuits, memory cells, input-output chips and interconnections. The effort foundered on the failure to develop a small, reliable high-speed "cache" memory (a dedicated memory for the central processor). Among other problems, the memory cells could not be packed too densely because magnetic flux in one cell would affect adjacent cells. At the same time that expectations for Josephson-junction memories were being downgraded, semiconductor memory chips were improving rapidly. In 1983 IBM terminated its project.

Japanese companies have continued to work on the Josephson computer. They improved the niobium-based Josephson junctions and are now developing components containing niobium nitride, which has a higher critical temperature and should yield faster circuits. They have not, however, demonstrated a reliable technology for manufacturing Josephson memory chips containing 1,000 or more cells. Memory circuits made of high-temperature superconductors will be even more difficult to produce, because the new materials appear to be trickier to process than niobium nitride.

A Josephson computer based entirely on high-temperature superconductors, then, appears to be infeasible. A hybrid system, however, combining Josephson logic circuits and semiconductor memory, all operating at 77 degrees K, may be viable. Indeed, gal-



MAGNETOENCEPHALOGRAPHY employs SQUIDs to map magnetic fields induced by the brain's electrical activity. For simplicity, the drawing shows an electric current (*gray arrow*) moving in one direction in the brain. The SQUIDs measure the perpendicular components of the magnetic field at the scalp; these components have a maximum and minimum value at the two points where the looping field lines penetrate the scalp at right angles (*red arrow*). The distance between the points depends on the current's depth; the midpoint between them lies on the same radius as the current.

lium-arsenide and metal-oxide semiconductor circuits perform better at 77 K. Such a hybrid would exploit the best features of both semiconductors and superconductors. The challenge is to develop materials and techniques to build Josephson junctions from the new superconductors and to design high-speed interfaces between the superconductors and semiconductors.

If fabrication problems could be overcome, the switching time of Josephson-junction circuits could be as little as 10^{-13} second for a superconductor with a transition temperature of 10 degrees K and 10^{-14} second for a material with a transition temperature of 100 K. No conventional technology is likely to approach these speeds. The devices could be incorporated into parallel-processing architectures for supercomputers 1,000 times faster than those now on the drawing board. It will take many years of research, however, to realize such a computer.

In the nearer term, superconductors may be limited to interconnections linking one semiconductor device to another. High-frequency signals traveling through superconducting wire undergo little attenuation (loss of energy) and no dispersion (spreading out). Such interconnections are most likely to be employed over distances of a few meters, for example to connect the components of a room-size computer. Over greater distances optical fibers are likely to retain an advantage in cost and performance. Superconductors are not likely to provide an advantage for the short interconnections on a chip either, where delay times are dictated primarily by the inherent impedance and capacitance of the device itself rather than by the resistance of the interconnections.

SQUIDs

Separate applications of Josephson junctions offer greater promise; indeed, some already exist. The superconducting quantum interference device, or SQUID, consists of one or two Josephson junctions inserted into a loop of superconducting wire. The device is extraordinarily sensitive to changes in electromagnetic fields. The magnetic flux contained within a superconducting loop is quantized: it occurs only in integer multiples of a basic unit called the fluxon (about 2×10^{-7} gauss centimeter squared). The current in the SQUID is a periodic function of the total fluxons contained by the loop and is exquisitely sensitive to any change in the magnetic flux.

A SQUID can measure voltage differences as small as 10^{-18} volt, currents as small as 10^{-18} ampere (only several electrons per second) and magnetic fields of less than 10^{-14} tesla (one ten-billionth of the earth's magnetic field). When a SQUID is operated at four degrees K, thermal noise is virtually eliminated and the SQUID's sensitivity approaches the fundamental limits imposed by quantum mechanics. No other technology comes close.

Physicists employ SQUIDs to search for exotic particles such as quarks, magnetic monopoles and gravitons and to test the predictions of general relativity. Geologists prospect with SQUIDs for oil, water and mineral deposits, which cause local anomalies in the earth's magnetic field. Such surveys are carried out from helicopters; the new superconductors may make possible more compact devices. On the other hand, there is more thermal noise at 77 degrees K, which might obscure weak signals. The Department of Defense is developing submarine detectors built of SQUIDs. In the future, SQUIDs may detect tiny ion currents in corroding metal.

SQUIDs are also beginning to find use to detect the faint magnetic signals associated with electrical activity in the heart and brain. Magnetoencephalography detects signals from the brain that are only slightly greater than 10^{-13} tesla and can determine the source of the nerve signal to within a few millimeters. Typically these magnetic signals decrease with the fourth power of the distance from the source, and so the pickup coils need to be as close to the head as possible—but today's devices, cooled with liquid helium, require thick insulation. Coils made of high-temperature superconductor cooled with liquid nitrogen would not need such thick insulation and so could be brought closer to the head. The rest of the apparatus would have to operate at four degrees K in order to achieve enough sensitivity.

Josephson junctions might also provide a new detector for electromagnetic radiation. At present there is no electronic device that can operate within a significant band of the infrared spectrum, between 10^{11} and 10^{13} hertz. The frequency to which conventional superconducting metals are sensitive increases with the critical temperature. If this relation also turns out to be true for the new superconductors, their higher transition temperature may pave the way for devices capable of detecting and transmitting this as yet unexploited part of the electromagnetic spectrum.

Josephson junctions have detected frequencies of up to 10^{11} hertz with efficiencies that approach the quantum limit (that is, they can detect almost every electron that has absorbed a photon). No conventional technology approaches this efficiency above 10^{10} hertz. The new superconductors may extend the performance to 10^{12} hertz. The first applications will probably be military ones, such as high-resolution radar, passive sensors and space communications.

Theoretical Frontiers

Physicists are able to describe much of the behavior of conventional superconductors by the BCS theory, named for its authors: John Bardeen, Leon N. Cooper and J. Robert Schrieffer of the University of Illinois. In normal conductors the conduction electrons lose energy as they scatter off impurities and vibrations, called phonons, in the crystal lattice. Each electron is assumed to distort the lattice locally and create a region of attractive force. In a superconductor below the transition temperature the force is thought to overcome the electrons' mutual electrostatic repulsion, enabling them to form bound pairs. What is more, all the electron pairs are in the same quantum state. It therefore becomes impossible for individual electrons to scatter off impurities (because all the other electrons would have to do so at the same time), and so the material becomes superconducting. The BCS theory showed that many features of the superconducting state, such as transition temperature, could be related to a few properties, including the strength of the assumed attraction.

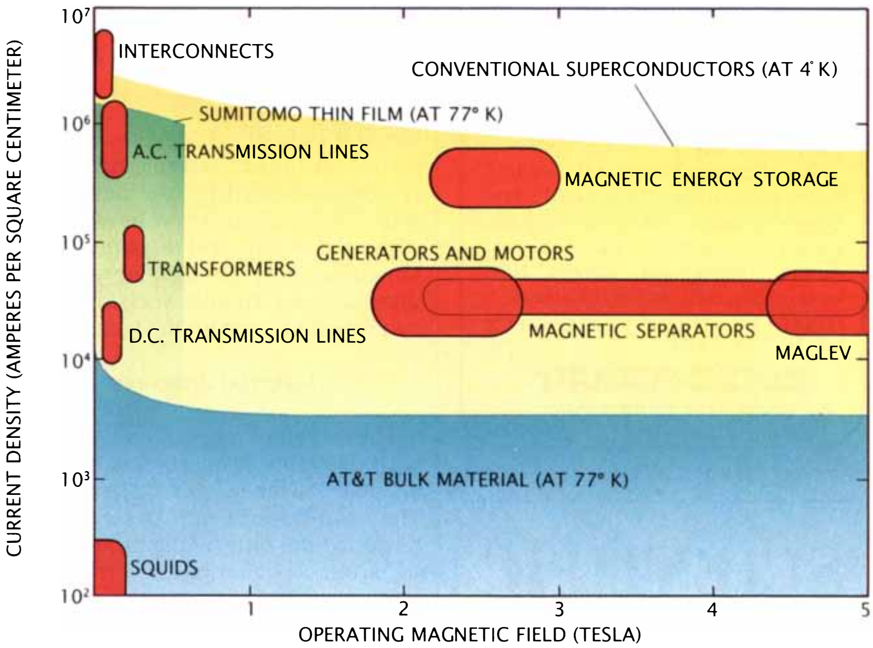
The new superconductors challenge this picture in two ways. First, the net attraction that would result from the lattice structure of the new superconductors does not appear to be strong enough to account for the high transition temperatures. Second, the role of the lattice has been called into question by some experiments that find no change in superconducting properties when normal oxygen in the lattice is replaced by different oxygen isotopes. According to the BCS theory, the change in oxygen mass changes the lattice inertia (and hence its vibration frequencies) and should therefore also change the superconducting properties. Other experiments, however, have observed this "isotope effect," and at the time of this writing the experimental situation and its interpretation are confused.

Electron pairs are expected to play a role in the new superconductors: experiments show that the smallest magnetic flux penetrating the materials corresponds to the presence of charge carriers with twice the electron's charge. Indeed, pairing is thought to be necessary for superconductivity because many bound pairs of electrons can be in the same quantum state, which allows the pairs to flow together and to exhibit some of the macroscopic effects associated with superconductivity. It has been difficult, however, to build a convincing case to support any particular pairing mechanism.

Theoretical understanding at the atomic level, however, may not be as essential to progress as some people think. Indeed, one should not expect to understand the new materials just by deducing macroscopic behavior from microscopic assumptions. The problem may simply be too complex to allow such deductions. Instead an artful mixture of macroscopic and intermediate-scale (10^{-8} meter) assumptions may be more fruitful, as V. L. Ginzburg, Lev D. Landau and A. A. Abrikosov of the Academy of Sciences of the U.S.S.R. and others have found.

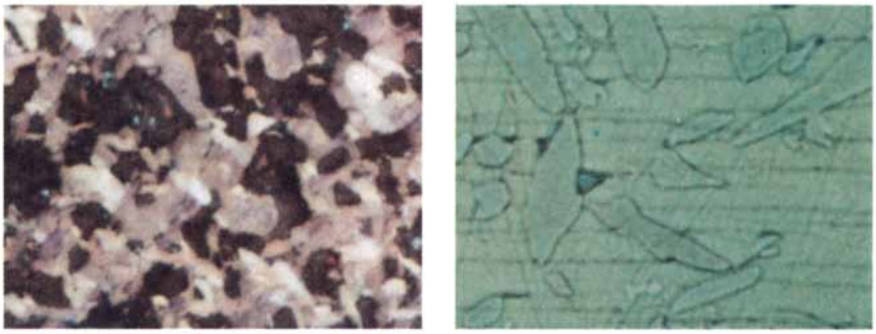
If theorists focus only on electron pairs and crystal lattices, they will neglect aspects of superconductivity that may be more crucial to practical applications. For example, imperfections in the lattice appear to stabilize the superconducting state by "pinning" large magnetic fields, thereby preventing them from inducing electric fields that would force the flowing current out of the superconducting state. It would be valuable to understand why that happens.

As is usually the case, theory will best be guided by experiment. In particular, it will be important to observe how time-dependent fields affect the new superconductors. Time-dependent currents and fields are unavoida-



DESIRED PERFORMANCE of superconductors based on present designs for devices are shown in this graph, along with the capabilities of existing materials. The materials must carry large enough electric currents and operate in strong magnetic fields. Niobium-titanium and niobium-3-tin operating at four degrees Kelvin meet the requirements for all applications except possibly SMES. The performance of high-temperature superconductors has improved markedly since their discovery. It must improve by one or two more orders of magnitude to satisfy most demands.

SUPERCONDUCTING TAPE AND WIRE have been fabricated by Roger B. Poeppel and his colleagues at the Argonne National Laboratory. Before sintering, the tape (*left*) is soft enough to be molded into various shapes; after sintering it becomes extremely brittle. A modest amount of flexibility has been achieved in a sintered coil (*right*).



MELT TEXTURING transforms disordered crystals of yttrium-barium-copper oxide (*left*) into long crystals (*right*). The horizontal hairlines are the boundaries of the melt-textured crystals. The irregular grains scattered over the surface do not impede electrical currents. The original sample carried only 400 amperes per square centimeter of electricity, whereas the melt-textured sample carries 17,000. The work was done by Sungho Jin and his collaborators at the AT&T Bell Laboratories.

ble and ubiquitous: devices must be turned on and off, and electrical power is often supplied by alternating current. One needs to know how the new superconductors will respond in such situations, how much energy they will dissipate and whether such losses can be remedied through design.

Material Progress

The design of new superconductors with desirable properties is more an art than a science, rather like cooking. The properties of puff pastry are determined not only by the ratio of flour to butter but also by the proper formation of layers. Similarly, the bulk properties of the new superconductors depend not only on their chemical composition but also on the alignment of individual crystals or grains. In particular, the electrical contact between two crystals appears to depend on how the crystals are oriented in relation to one another.

The microscopic features reflect the macroscopic methods by which the material is made. The standard "shake and bake" method grinds together several metal-oxide powders, compresses the mixture and then sinters (bakes) it in an oxygen atmosphere at about 900 degrees C. The elements in the powders react chemically to form the new compound and fuse into a solid. The process also produces partially reacted compounds that coat the superconducting crystals, insulating them somewhat from one another; moreover, the crystals are not aligned to make good electrical contact. One remedy is "melt texturing": melt the sintered product and then freeze it. The resulting material has better crystal alignment, and the intergranular debris is pushed aside. Other preparation methods are being explored.

The new materials must meet a formidable array of physical requirements, depending on the applications. They must be able to carry currents of more than 100,000 amperes per square centimeter at 77 degrees K. For many applications the superconductors must be strong and flexible enough to withstand large magnetic, gravitational and centrifugal forces (in spinning turbine generators and motors, for example). In addition they must be easy to shape. Although niobium-3-tin is electrically and magnetically superior to niobium-titanium, the latter finds more application because it is easier to work.

Another problem is that some of the new compounds are chemically unstable: yttrium-barium-copper oxide gives up oxygen to surrounding materials, losing its superconductivity. New composites must be developed to ensure stability. For example, one might sheathe the material with silver, which does not pick up oxygen from the superconductor, and then surround the silver with less expensive copper. The task will not be simple: making a composite requires consummate craft. Previous success in making composites of niobium-titanium, copper and nickel gives hope that analogous feats can be achieved with the new superconductors.

Fabricating useful products—bulk items, wire, tape and thin film—is a formidable challenge. Thin films for electronic devices are made by depositing a gossamer layer of the component elements onto a substrate such as magnesium oxide. The chemicals are deposited as vapor, as evaporation from heated chemicals or as a laser-vaporized blast. Afterward the thin film usually must be annealed (repeatedly heated and slowly cooled) to provide the proper oxygen content. Every variable in the process can affect the outcome. Thin-film fabrication processes result in cleaner and better-aligned material. Many laboratories have made thin films that can carry one million amperes per square centimeter. Some people are considering whether thin films can be attached to a backing material and wound like a roll of paper towels to make a conductor capable of carrying large currents.

Tape and wire have been fabricated by molding a plastic mixture of superconducting powder. The mixture is either spread on a flat surface to make tape or extruded through a die to make wire. The material is then sintered and annealed. This technique has yielded wire capable of carrying current at 77 degrees K. Unfortunately the wire is brittle. The hope is that better properties will be attained as they have been with glass, which can be as brittle as a window pane or as supple as optical fiber.

These challenges may be as difficult as the discovery of the new superconductors was. Making practical materials will require sustained effort for at least a decade: time for experienced scientists to shift their attention to the new enterprise and time to train young scientists. Leaders of industrial research in Japan appear willing to make such long-term commitments. Four U.S. companies, AT&T Bell Laboratories, Bellcore, Du Pont and IBM, conduct sizable research programs. More people and resources, now dispersed among the private sector, national laboratories and universities, will have to contribute knowledge and skills from diverse disciplines—physics, chemistry, electrical engineering and materials science—if the high-temperature superconductors are soon to live up to their promise.

FURTHER READING

APPLICATIONS OF HIGH TEMPERATURE SUPERCONDUCTIVITY. A. P. Malozemoff, W. J. Gallagher and R. E. Schwall in *Chemistry of High Temperature Superconductors,* edited by D. L. Nelson, M. S. Wittingham and T. F. George. American Chemical Society, 1987.

HIGH TC SUPERCONDUCTORS AND POTENTIAL APPLICATIONS. Commission of the European Communities, 200 Rue de la Loi, b-1049, Bruxelles; July, 1987.

ADVANCES IN APPLIED SUPERCONDUCTIVITY: A PRELIMINARY EVALUATION OF GOALS AND IMPACTS. A. M. Wolsky et al. Argonne National Laboratory, Argonne, Ill. 60439; 1988.

SUPERCONDUCTIVITY RESEARCH: A DIFFERENT VIEW. John M. Rowell in *Physics Today,* Vol. 41, No. 11, pages 38–46; November, 1988.

SCIENCE AND BUSINESS

Fast computers, can-do steel, model mice

Supercomputing

Preeminence in computing takes more than a fast machine

It used to be that anyone who wanted to buy the fastest computers in the world trekked to Chippewa Falls, Wis., and chatted with Seymour Cray. A legend in his own time, Cray still designs and builds supercomputers, multimillion-dollar machines that solve complex numerical problems faster than any other computers. By 1986, a decade after its first supercomputer left Cray Research's Wisconsin workshop, the company was earning a tidy profit—almost $125 million that year—and had installed more than 120 machines, many in U.S. Government research laboratories.

Cray Research still sells 80 percent of the supercomputers outside Japan. But the company faces growing competition, much of it from Japanese companies that are producing machines with processing speeds that rival the speeds of the fastest CRAY's. And Cray, Control Data (which also makes supercomputers) and others hoping to enter the marketplace will have to work hard to match the sheer muscle of the huge Japanese electronics companies, which make their own components, are savvy in advanced manufacturing techniques and have the financial stamina to endure the ups and downs of demand for supercomputers. Only IBM, which is also reentering the supercomputer field, has comparable clout.

Moreover, except for a handful of aerospace, petroleum and automobile companies, U.S. industry has approached supercomputers with more caution than enthusiasm. Potential buyers now insist on ample software, easier-to-use systems and data-communications networks to link multiple investigators to supercomputers. "We [Cray] have to let go of the old goals and images of just making the fastest hardware and doing it out of the way, in the woods," says John A. Rollwagen, chairman of Cray Research. "The fastest hardware is a necessary but not sufficient condition for success."

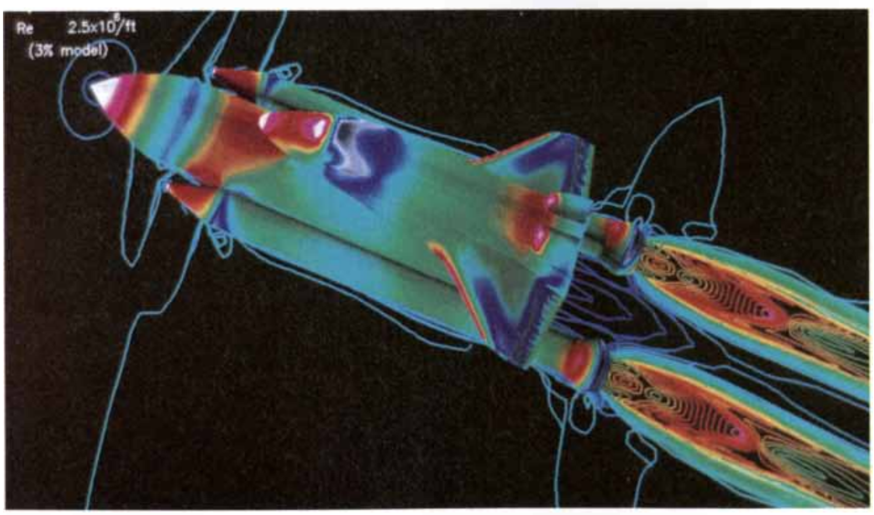
Building the fastest machines demands the most sophisticated hardware and software. High-performance memory chips, processors and other components must be packed tightly to trim the distances signals travel; the algorithms dictating how the machine operates and the software for applications must be written efficiently.

So far U.S. makers, led by Cray, and their Japanese competitors have followed different paths in developing hardware. U.S. designers are increasing the speeds of their machines by tying together multiple processors, which operate simultaneously on different parts of a problem. The CRAY-3, for example, due to be delivered to customers in 1990, will employ as many as 16 processors to reach peak processing speeds of 16 billion floating-point operations per second. To take full advantage of such speeds, however, users must learn to break up problems in order to run efficiently on multiple processors. Current Japanese supercomputers, in contrast, rely on one "vector" processor, which rapidly performs the same set of operations on a long sequence of numbers, teamed with software that rearranges code to use the processor efficiently. (U.S. companies use vector processing less extensively.)

Approaches to software differ too. U.S. makers rely increasingly on operating systems based on the widely used UNIX software. Much applications software is based on UNIX, which makes it possible (although not simple) to run an application on a variety of machines. Fujitsu, NEC and Hitachi, on the other hand, have built supercomputers with proprietary operating systems, compatible with the mainframes they make. The policy helps their customers to move smoothly from mainframe computing to supercomputing, explains Kenichi Miura, a vice-president at Fujitsu America. But it has also meant that the Japanese machines cannot run UNIX-based applications. Now Japanese producers are moving to UNIX-based operating systems.

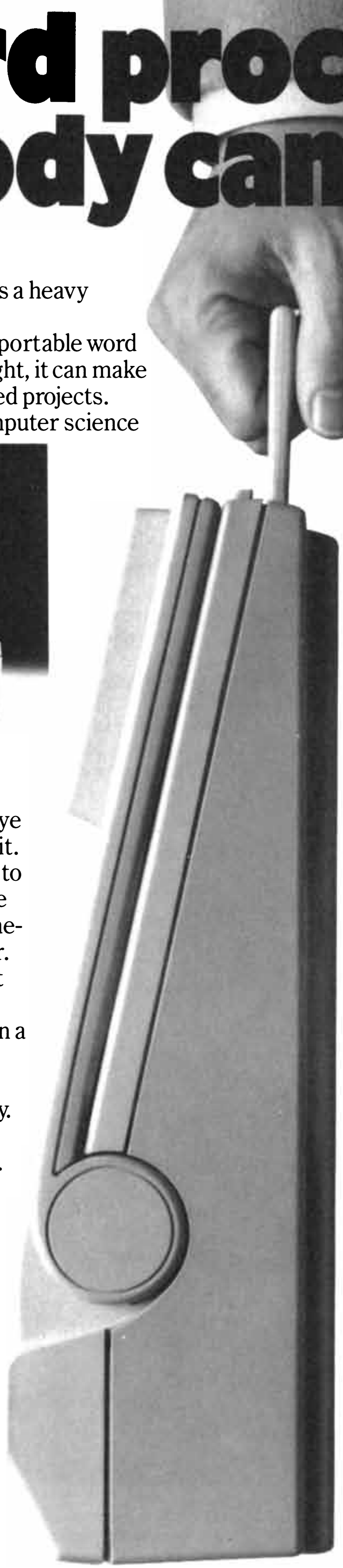
Japanese supercomputers work with a more limited range of numbers than do CRAY's, albeit they represent those numbers with greater precision. The ability to cope with very large numbers becomes increasingly critical as problems become more complex.

The diversified nature of the Japanese companies building supercomputers means that they make all the components needed for their machines. Cray and other U.S. companies must rely on alliances and independent suppliers—including their Japa-



***SURFACE AIR PRESSURE** on the space shuttle less than a minute after launch is calculated on a CRAY-2 supercomputer at the National Aeronautics and Space Administration's Ames Research Center at Moffett Field, Calif. The calculations, based on about a million data points, took 20 hours to complete.*

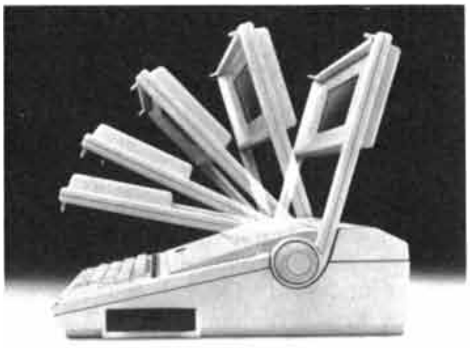
Word processing anybody can pick up.



Who says word processing is a heavy subject?

The Smith Corona PWP 80 portable word processor is not only light in weight, it can make light work of the most complicated projects.

You don't need a PhD in computer science to understand it.



Our screen adjusts to you, so you don't have to adjust to it.

You don't have to keep one eye on the instruction manual to use it.

In fact, PWP 80 is so simple to operate, you may not even believe you're actually using a state-of-the-art, fully featured word processor.

All you have to do is unfold it and begin.

You type just as you would on a typewriter. Your text appears on the backlit, supertwist 16 line by 80 character liquid crystal display.

Block moves are a cinch, inserts a snap, deletions a breeze.

When you're ready to print, just insert a sheet of paper, select Print and presto, letter-quality printing.

And that's it.

With PWP 80, everything is at your fingertips.

If you're only writing something short, like a note or recipe, PWP 80 can also function as a conventional, state-of-the-art typewriter.

Of course, there's nothing conventional about PWP 80's features.

Inside, there's 50,000 characters of internal editable memory. (That's about 25 pages.) There's also a built-in disk drive, which makes storage capacity virtually unlimited.

With our amazing Grammar-Right System,™ you'll never write a wrong again. It includes a 75,000 word Spell-Right™ dictionary that alerts you to misspelled words, an Electronic Thesaurus that gives you alternatives to over-used words, and Phrase Alert,™ which points out inappropriate phrases—before anyone else does.

DataDisks give you unlimited storage capacity.

And for those of you who are unfamiliar with word processors, we've even included a special Tutorial DataDisk, so you can learn how to use PWP 80 in practically no time.

Best of all, this giant among word processors weighs under 20 pounds. Which means PWP 80 can go wherever you go. From home to school to office.

And that might just be its best feature of all. Because when you consider how remarkable PWP 80 is, it's a wonder you'll want to go anywhere without it.

SMITH CORONA®

TOMORROW'S TECHNOLOGY AT YOUR TOUCH

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

nese competitors—for hardware components. Steve S. Chen, formerly a top designer at Cray Research who started his own supercomputer company in 1987 and has since received backing from IBM, has complained that many of the critical components he needs are not made by U.S. companies.

Rollwagen notes that so far Cray has found Fujitsu to be a reliable components supplier; the semiconductor division has told him it often prefers to test new chips with Cray rather than with Fujitsu's own computer divisions. If Japanese suppliers did want to favor their own producers, "they have easy leverage over us," he acknowledges. "But frankly we have no alternative."

Lack of advanced manufacturing skills has been another hurdle for Cray and may well become one for smaller U.S. companies. Processor modules in the CRAY-3, for instance, will measure 16 square inches and will be attached to other modules by 12,000 connectors; only a robotic assembly line can manufacture the systems precisely. Difficulties in setting up the production line helped to delay the release of the CRAY-3 by more than six months.

Besides disappointing customers, product delays are sharply noted by Wall Street. In late October, Cray stock dropped about 17 percent as analysts learned not only of a decline in quarterly revenues but also of manufacturing delays and problems in developing the CRAY-3 logic chips, which meant the company had to spend an additional $10 million on development.

Even as Cray and other U.S. makers wrestle with such challenges, many supercomputer users wonder if it really matters which company—or country—builds the fastest computers. Sales will not swing the trade deficit. In 1988 worldwide sales of supercomputers edged near $1 billion—less than a fourth of the annual U.S. sales of movie tickets. "The benefits of supercomputing do not come from the creation of the machines, they come from the use of the machines," said Tennessee Senator Albert Gore, Jr., at a symposium held last fall by the National Academy of Sciences.

The sheer processing speed of supercomputers, more than 1,000 times that of a powerful personal computer, enables workers to tackle complex problems they would otherwise oversimplify or avoid, including modeling the airflow over the wing of a plane, ferreting out hidden oil reserves based on reams of seismic data or mapping the human genome. Moreover, the machines can turn a flood of data into a three-dimensional, color-coded image, which can spark investigators to ask new questions. Supercomputers are "tools of the imagination," says Sidney Karin, director of the supercomputing center at the University of California at San Diego. Provided that workers have access to the fastest machines, Karin adds, it does not matter where they are made.

"People said we didn't need to make consumer electronics or DRAM's [dynamic random-access memory chips]," retorts Alan K. McAdams, an economist at the Johnson Graduate School of Management at Cornell University; now observers mourn the loss of those industries. Last summer McAdams helped to write a report for the Institute of Electrical and Electronics Engineers arguing that a strong domestic supercomputer industry is essential if the U.S. is to become a prime user of supercomputers.

Spurring U.S. companies to buy supercomputers may be the greatest challenge, regardless of who builds them. Rollwagen notes that of the 50 orders for supercomputers Cray won in 11 months of 1988, only four or five came from U.S. companies; universities bought 18, the Government 15 and foreign companies a dozen.

Since the hardware alone for a high-end supercomputer can cost $20 million, the decision to buy one is not made lightly. "But our biggest obstacle is finding ways to assimilate the use of supercomputers," says Clifford R. Perry, a research director at the Eastman Kodak Research Laboratories. Intimidated by supercomputers, investigators may prefer to continue working with computers they already understand, he says. Partnerships with university supercomputer centers can be a good way for corporate workers to learn about the machines, says Edward A. Mason, a vice-president at Amoco. More important, however, will be the day software runs seamlessly on both a desktop computer and a supercomputer; then scattered workers connected by a network will be able to tap the power of an anonymous supercomputer. *—Elizabeth Corcoran*

STEEL FROM INTEGRATED MILLS is used to make about three billion beverage cans annually in the U.S., less than 5 percent of the total. Steelmakers are now hoping to boost their share of that market.

Sharpening the Edge

Will steel beverage cans return to store shelves?

There is a battle heating up over the cans in the cooler. In the 1960's almost all beer and soda cans were made from tin-plated steel; about 95 percent of the 80 billion cans made last year in the U.S. were aluminum. Within the past two years, however, improved steel manufacturing processes and rising aluminum prices have encouraged steel companies to challenge aluminum's command of the beverage-can market.

The typical steel beverage can of yesteryear was a heavy, leak-prone contraption, welded together from three thick parts that flavored the drink with a metallic taste. Then aluminum producers developed a process for drawing a can of aluminum from a flat sheet; the can lacked only a lid. The primary objection to aluminum cans was that they would not rust away when buried in dumps. Consequently the aluminum industry mounted a large recycling campaign as it wooed can makers to aluminum.

Within the past five or six years, however, the steel companies that survived the decline of the industry began improving their plants. Mills once began with steel ingots, rolled them into .009-inch-thick sheets and then plated the sheets with tin. Now the same manufacturers use a continuous-casting process that produces long ribbons of hot steel, which are later squeezed down to .003 inch. The thickness of the tin coating has been reduced by half. The thinner steel can be punched into seamless cans much as aluminum is. Such steel cans still have aluminum lids; there is as yet no such thing as a steel flip-top. Still, steelmakers argue that the rising price of aluminum means a beverage-can maker can save from about $4 to $6 per lot of 1,000 cans by using steel instead of aluminum.

Few can makers have switched to steel cans yet. Assembly lines must be retooled; refitting a dedicated aluminum line to make steel cans may cost as much as $3 million. (Lines built to manufacture both steel and aluminum cans might require only $200,000 to return to steel, says Charles G. Carson, a manager at U.S. Steel.)

An even greater hurdle for the steelmakers is building a nationwide recycling program comparable to one that recycles roughly half of the aluminum beverage cans in the U.S. The public now expects beverage cans to be recycled, steelmakers say, but the economics of recycling work against steel. Because the thin sheet steel used in cans is a high-quality steel, producers must carefully monitor the level of other metals—tin in particular—in recycled steel to ensure it does not become brittle. "In many cases it's more expensive to recycle [a steel can] than to produce one from scratch," says John E. Jacobson, a vice-president at AUS, a consulting firm in Philadelphia.

Even though the minimills, the fastest-growing segment of the steel industry, depend on scrap steel, they have lacked the heavy rolling equipment needed to make steel plate. Within months, however, Nucor is to open a plant in Indiana that will use a continuous-casting technique pioneered in West Germany to make flat-sheet steel. Nucor will first produce .25- to .1-inch-thick sheets, but it hopes eventually to refine the process and produce high-quality thin sheets.

In quest of a broad-based steel-can recycling program, the six U.S. producers of tin-plated steel have said they will pool some $30 million over the next five years to sponsor an institute for steel-can recycling. The organization will develop local recycling networks, taking a cue from a successful steel recycling program in Ontario, and publicize the programs.

Meanwhile makers of aluminum cans have not been idle. Although only 10 percent of food cans are now made from aluminum, Alcoa is aggressively marketing its small but easy-to-open cans for food. *—E.C.*

A Tiny Mouse Came Forth

New product release: Du Pont's patented oncomouse

There is a new mouse at the Charles River mouse farm this year. In fact, there soon will be several hundred genetically identical new mice down at the farm, which raises rodents for clients. They are "OncoMice," made, in a sense, by Du Pont, the erstwhile chemical company.

The oncomice are the creatures that jumped out of the annals of science and onto the front page of newspapers around the country last April, when the U.S. Patent and Trademark Office granted Harvard University the first patent for a transgenic animal. Such mice, genetically predisposed to developing cancer, are expected to aid workers testing new therapies and—because they are better models of the disease—cut down the number of animals used in experimentation.

The new breed was developed by Harvard workers Philip Leder and Timothy A. Stewart, who inserted a human oncogene into a mouse embryo soon after conception. Since much of the funding for Leder's research was provided by Du Pont, the chemical company had commercial rights to the results of the work.

Still, Du Pont did not expect a patented mouse to emerge from the contract, says Andrew S. Forman, who became Du Pont's oncomouse product manager. "But the day I learned of the patent," he recalls, "I said, 'How do we turn this into a business for us?'"

It took some six months and an earnest sales pitch to convince Du Pont's management that mice would fit nicely into the company's product lines. That done, the company obtained four mice from Leder's laboratory and began breeding.

"Leder invents, we rederive," Forman says. Du Pont began with heterozygous oncomice (mice that carry one copy of the oncogene each) and crossbred the rodents to produce a stud pool of verified homozygous mice—ones carrying two copies of the oncogene. Charles River takes over in January. To produce mice for sale, it is mating the inbred strain of homozygous male oncomice with a similarly inbred strain of normal females. The intensive inbreeding results in genetically identical, heterozygous pups.

At Charles River the mice lodge in sterilized rooms monitored for contaminants by computers. The company even preserves some mice strains by freezing embryos in liquid nitrogen—just in case a line spontaneously mutates. When the mice leave the farm, they travel in special crates, hauled by climate-controlled trucks.

Du Pont will offer the mice initially to nonprofit research groups and universities. The price? About $50 per mouse, with discounts for large orders. Charles River says the price is competitive with other specialty mice; commoner varieties may run as low as $1 per animal.

There is, Forman concedes, nothing to stop investigators from breeding oncomice in their own laboratories. Although Du Pont considered selling animals of only one sex or sterilizing the mice as ways to protect its patent rights, it decided to rely on the honor system and on the fact that workers often need to be certain their mice are genetically uniform.

Since the patent technically covers all transgenic nonhuman mammals that carry oncogenes and there are about 40 known oncogenes, Forman predicts that the present mouse may well be only the first for Du Pont. He says the company plans to write to investigators who have made their own oncomice requesting a nominal user's license fee and asking whether they would like to collaborate with Du Pont's mouse project. Nevertheless, Forman adds, "we don't want to impede research." *—E.C.*

From Bird Song to Neurogenesis

Studies of song-control centers in the canary brain reveal that new nerve cells are born in adulthood and that they can replace older cells. Such neurogenesis could hold the key to brain self-repair in humans

by Fernando Nottebohm

One of the most firmly established beliefs in the neurosciences has long been that all neurons, or nerve cells, in the brain of vertebrates are formed early in development, when the brain itself is growing. An adult vertebrate, it was believed, must make do with a fixed number of neurons. Hence neurons lost through disease or injury are not replaced, and learning takes place not by putting new cells into the neural circuits that control behavior but by modifying connections among a limited number of neurons.

Joseph Altman of Purdue University first challenged this belief in the early 1960's on the basis of experiments he did on rats and cats. He argued that his experimental results showed that some types of neurons continued to be formed in parts of the animals' brain even after they had reached adulthood. Altman's results, however, were not conclusive, and neurogenesis (the birth of neurons) in the adult mammalian brain is still not generally accepted.

New evidence for such neurogenesis in another class of vertebrates has now come from a most unexpected quarter: the study of how birds learn to sing. Recent experiments carried out by my colleagues and me show not only that neurons are constantly being born in the brain of birds after the birds reach maturity but also that the newborn neurons can in some cases replace older ones. Although we have not yet shown exactly what the new neurons do, we think they are used to acquire new information. Song learning by both juvenile and adult birds may therefore depend on the availability of young neurons that can be used to build novel circuits. These findings raise questions about the general stability of the neural circuits in the brain. Perhaps their most tantalizing aspect, however, is that they might eventually lead to the identification of factors that could stimulate a human brain to repair itself by replacing damaged neurons with new ones.

FERNANDO NOTTEBOHM is professor of animal behavior at Rockefeller University and director of the university's Field Research Center for Ethology and Ecology. A native of Argentina, he originally came to the U.S. to study agriculture but earned a B.A. and Ph.D. in zoology from the University of California, Berkeley. He joined the faculty at Rockefeller as assistant professor in 1967 and was made full professor in 1976.

Taxonomists recognize some 30 different orders of birds, containing a total of about 8,500 living species. Almost half of the species are classified into the songbird suborder, Oscines, of the order Passeriformes. Songbirds tend to stand out among other types of birds because of their rich and varied song. A bird sings to announce its presence to its neighbors and to claim a breeding territory. In addition birds (generally males) sing to attract mates.

It has long been common knowledge that some birds are able to imitate the sounds they hear. Yet before the 1950's few biologists realized this talent is routinely exercised by songbirds to develop their everyday song. W. H. Thorpe of the University of Cambridge first demonstrated this fact (and thereby established a new field of research) when he described how a European songbird, the chaffinch, learned its song. He reared isolated male chaffinches in soundproof chambers equipped with speakers. In some chambers he broadcast recorded chaffinch songs, which the young chaffinches were then able to imitate. Birds that had not been exposed to the recorded songs, however, developed abnormally simple songs. Moreover, exposing the deprived birds to the "tutor" tapes after they had reached sexual maturity did not improve their singing ability.

Thorpe concluded that birds learn to sing much as humans learn to speak, that is, by imitating models provided by their elders. He also concluded that song learning for birds such as the chaffinch is limited strictly to a "critical period" sometime before sexual maturity. Subsequent work by Peter R. Marler of Rockefeller University and Klaus Immelmann of the University of Bielefeld in West Germany showed that two other songbirds, the North American white-crowned sparrow and the Australian zebra finch, are also critical-period learners. Yet not all songbirds fall into that category. Canaries, for example, can alter their song from year to year; they are called open-ended learners.

The first sounds a newly hatched canary makes are shrill and high-pitched calls that spur its parents to feed it. Such "food begging" continues even after the juvenile has left the nest; it lasts until the bird becomes completely independent of its parents at about four weeks. Thereafter the bird produces its first rudimentary attempts at singing, which are called subsong. Subsong is soft in volume and variable in structure, and it is often done while the young bird appears to doze. Charles Darwin pointed out the similarity between subsong and the babbling of human infants; both seem to be early stages of vocal practice from which emerges the full repertoire of sounds used in communication.

As subsong becomes more structured by the end of the bird's second month, it is given a new name: plastic song. Plastic song sounds a bit like the song of adult canaries but is still quite variable. It becomes increasingly stereotyped as the bird approaches sexual maturity, at about seven or eight months. Canary fanciers have long known that the quality of a young bird's song is influenced at this stage by that of its older companions.

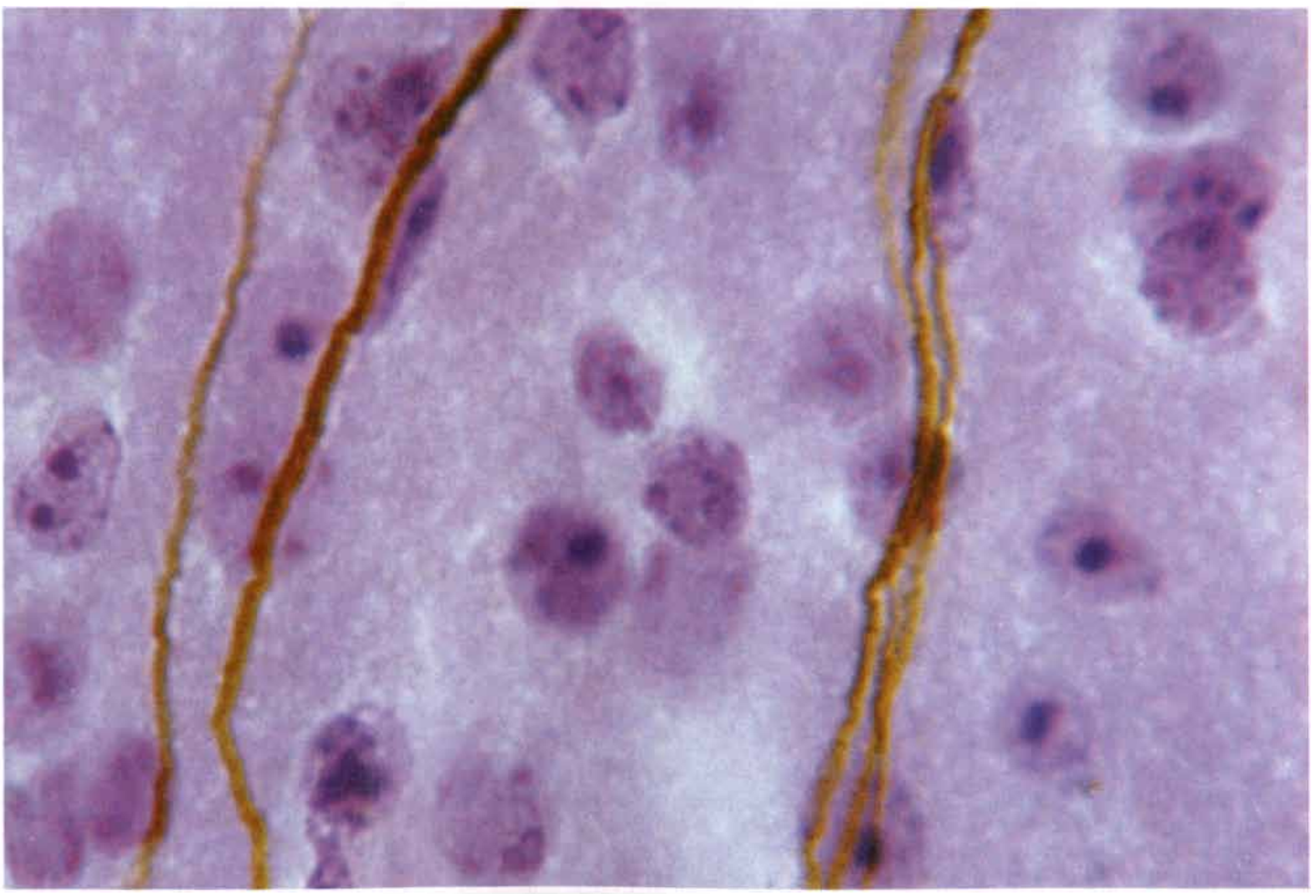
The final song pattern of an adult male canary, known as stable song, is sung for the duration of the bird's first breeding season. Such song can be characterized by the number of different sounds (called syllables) it contains. Although a male canary only three or four months old can already vocalize 90 percent of the syllables it will employ as an adult, it is not until sexual maturity that syllables become stereotyped. Indeed, achievement of stereotypy seems to be a difficult task for canaries, since it takes them several months of practice while in the plastic-song phase.

Yet even after such practice a canary's repertoire of syllables is not permanently learned. Every year during late summer and fall (after the breeding season) the syllable stereotypy mastered a few months earlier is lost, and the bird's song becomes as unstable as the plastic song of juvenile birds. In fact, many of the learned syllables disappear from its basic song "vocabulary" and new ones are acquired, which may then be incorporated into a stereotyped song the following winter and spring. In this way adult male canaries can develop a new song repertoire every year. Such seasonal song learning is probably under hormonal influence, since peaks of new syllable addition are preceded by drops in the level of the male sex hormone testosterone in the blood.

When Thorpe did his experiments with chaffinches, nothing was known about the parts of the avian brain that control song learning. It was not until 1976 that several anatomically distinct clusters of cells that control canary song were identified in my laboratory at Rockefeller University. Such a cluster of cells is called a nucleus. (It should not be confused with the nucleus of an individual cell, which contains the cell's genetic material.)

The largest of these nuclei, the higher vocal center (HVC), lies in the bird's forebrain. The axons—the long processes characteristic of neurons—of many of the cells in the HVC extend to another forebrain nucleus, the robustus archistriatalis (RA). Many RA neurons in turn have axons that contact hypoglossal motor neurons innervating the muscles of the syrinx, the organ that actually produces song [*see illustration on next page*]. Since these various brain nuclei are rather distinct, it is possible to estimate their volume accurately and relate it to a particular bird's sex and age, the hormone level in its blood and the complexity of its song. Although the brain of a young canary reaches adult size sometime between 15 and 30 days after hatching (at about the time the bird becomes independent of its parents), the HVC and the RA continue to grow for several more months—almost until the bird reaches sexual maturity. It is during this period of HVC and RA growth that young birds first learn to sing.

In 1976 Arthur P. Arnold (who was then at Rockefeller) and I discovered that the HVC and the RA were some three to four times larger in adult male canaries, which sing complex songs, than in adult female canaries, which sing simpler songs. It appeared that the amount of brain volume devoted to a particular skill was considerably greater in the sex excelling at that skill. This instance of so-called sexual dimorphism disproved another long-held view, namely that brains of



NEWBORN NEURONS adopt a characteristic elongated form as they migrate through the canary forebrain. Initially the neurons follow the long fibers (*brown*) of radial glia, a common type of cell in the avian brain. Both old and new brain cells have been made visible by staining their cell nuclei purple; two migrating cells can be seen, each on a different fiber.

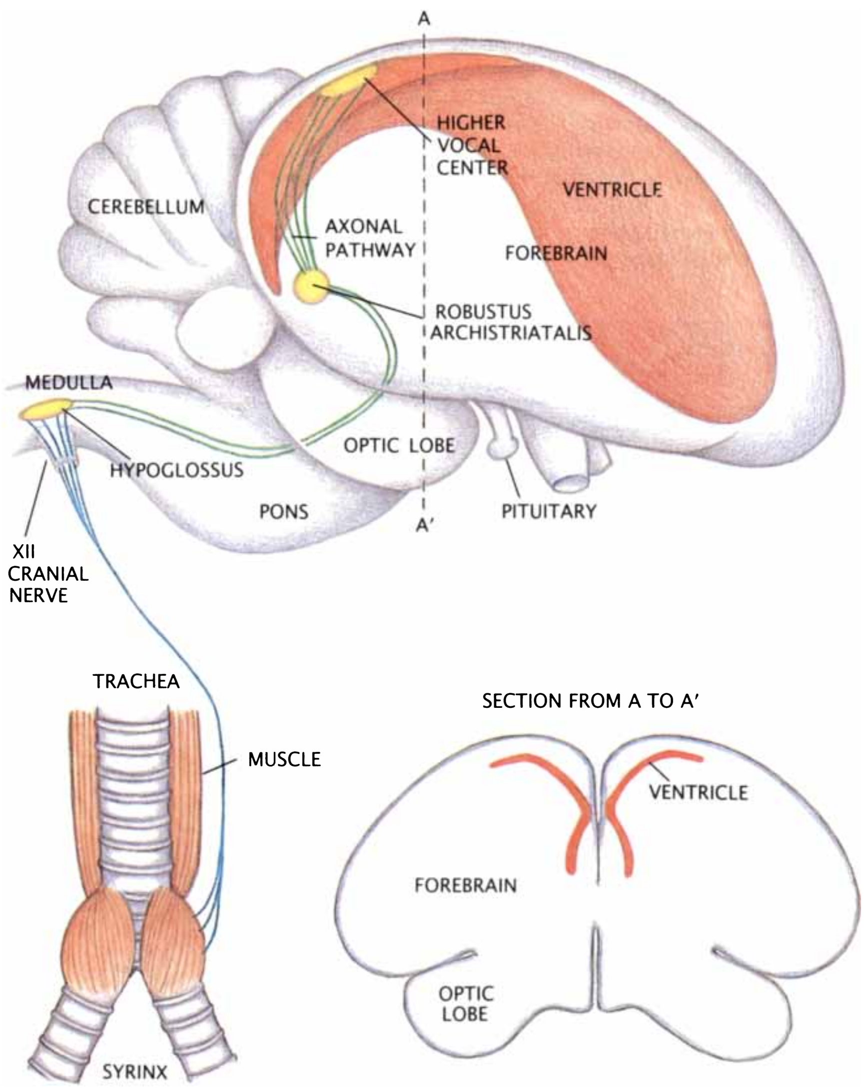
vertebrates exhibited no marked anatomical differences between the sexes.

Subsequent work by Mark Gurney and Masakazu Konishi of the California Institute of Technology showed that the sexual dimorphism of the RA arises (at least in part) from differences in the number of neurons it contains and that these differences become apparent early in development—even before song learning begins. These results suggest how brain anatomy might limit learning: the greater the number (and perhaps diversity) of neurons integrated into a particular neural circuit, the greater the amount of information the circuit can handle.

The relation between the size of the song-control nuclei and singing skill also holds for canaries of the same sex. Some male canaries are unusually talented singers and have developed a large repertoire of song syllables; they tend to have large HVC's and RA's. Other birds of the same breed, sex and age that have been kept under identical conditions produce simpler songs consisting of a smaller number of syllable types; they tend to have small HVC's and RA's.

Although it is tempting to infer that the anatomical differences are the basis of these marked differences in singing talent, there also are canaries that have a large HVC and nonetheless have a small syllable repertoire. An analogy between the size of the song-learning nuclei in canaries and the amount of shelf space in a library may help to explain the relation. If a library is to hold many books, it needs ample shelf space, yet the shelves of a large library need not be completely filled. (Actually, under some conditions the arrival of new "books" can expand the "shelf space." Observations by Sarah Bottjer of the University of Southern California and Arnold, who is now at the University of California at Los Angeles, suggest that the act of song learning itself can enlarge the size of the HVC.)

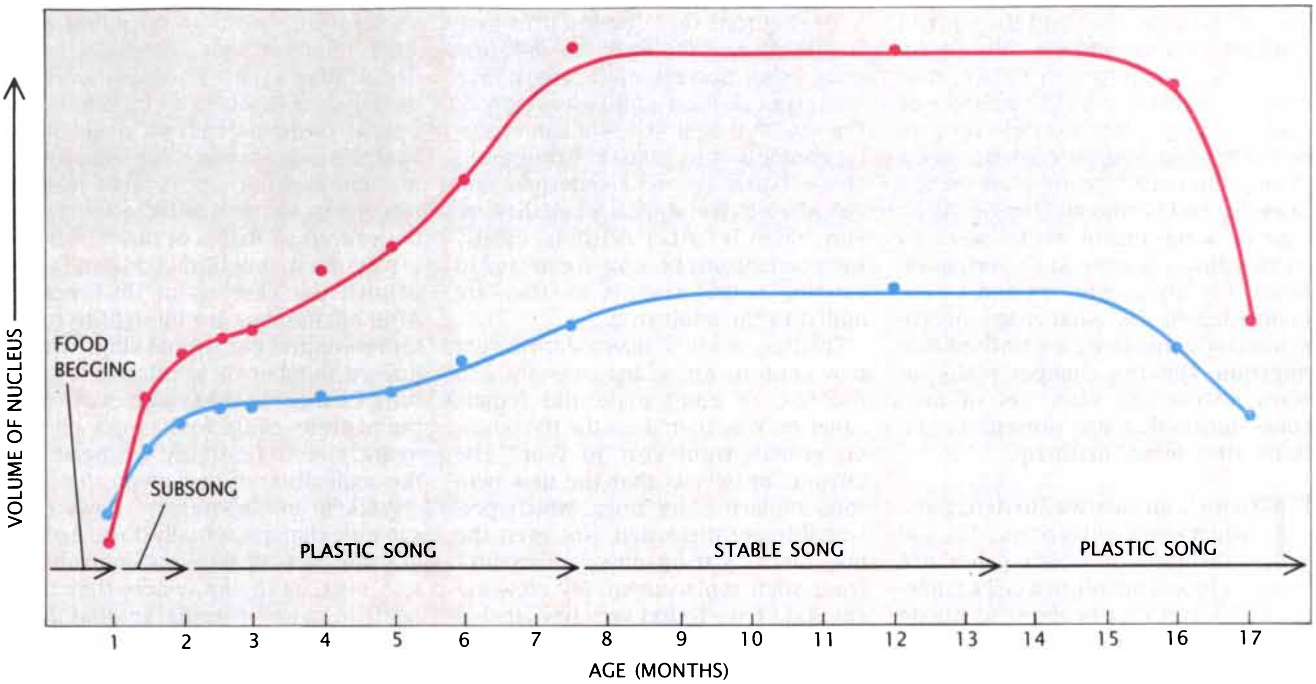
Statistical analyses of correlations between the size of syllable repertoires and the size of the HVC in adult male canaries show that only 20 percent of the variability in either one can be explained as a consequence of the other. Song learning in canaries, it appears, is also influenced by factors other than HVC and RA size. Nevertheless, observations made in my laboratory and elsewhere suggest that the number of neurons in the HVC and RA (as reflected in the physical size of these nuclei) does influence the skill with which a particular canary sings.



CANARY BRAIN, shown in a side view (*top*) and cross section (*lower right*), contains several nuclei, or distinct clusters of cells, that control song learning. The largest of these nuclei is the higher vocal center (HVC). Electrochemical signals from the HVC are transmitted to other parts of the brain along axons, the long processes characteristic of neurons, or nerve cells. Many HVC neurons have axons that extend to neurons in another song-control nucleus, the robustus archistriatalis (RA). The axons of many RA neurons in turn contact motor neurons of the hypoglossal nucleus that innervate the muscles of the syrinx, the organ where sound is actually produced.

Further evidence for the importance of the size of the HVC and RA in determining singing ability comes from the effects of testosterone. It is possible, for example, to elicit malelike song in a nonsinging adult female canary by giving it intramuscular injections of testosterone. The hormone does not merely activate existing neural circuits but doubles the volume of the HVC and RA in the female canary's brain. Similarly, in adult male canaries blood testosterone levels are very high in the spring, when their song is quite stereotyped, and low in early fall, when their song is as variable as that of juveniles. At the same time the HVC and RA in the spring are roughly twice as large as they are in the fall.

In 1981 Timothy DeVoogd, who was then at Rockefeller, and I studied how testosterone induces growth in the RA of adult birds (there are still no data on how the hormone enlarges the HVC). The most abundant neurons in the RA send out long axons that connect with hypoglossal motor neurons innervating the syrinx. Sprouting from the main body of these neurons are many secondary branches, called dendrites, that are consistently longer in males than they are in females. Yet in the female canaries we injected with testosterone these dendrites grew and became indistinguishable from those of males. As the dendrites grew they

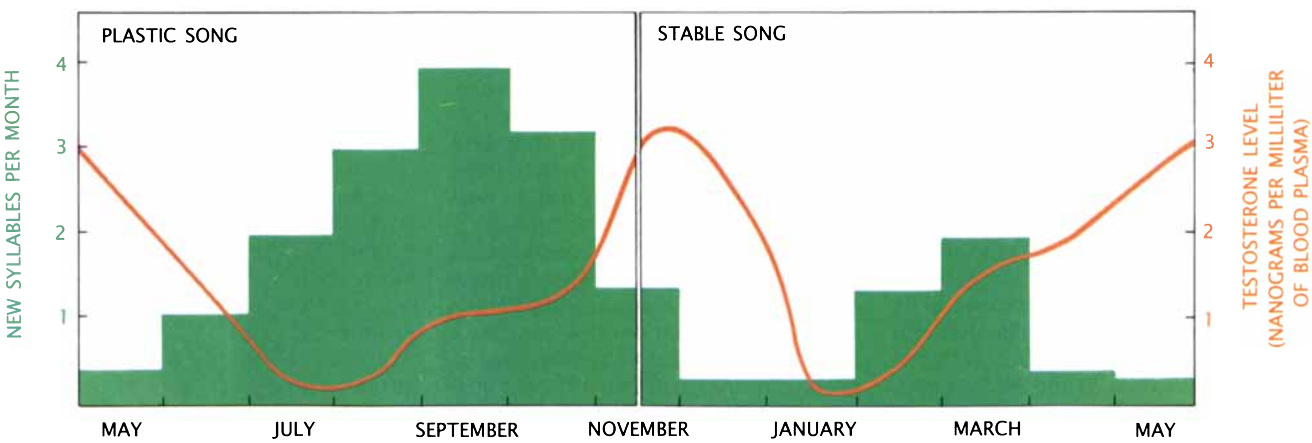


DEVELOPMENT OF SONG in male canaries is accompanied by a marked increase in the volume of both the HVC (*red*) and the RA (*blue*). Canary song develops through four different phases: food begging, subsong, plastic song and stable song. Food begging consists of shrill and high-pitched calls; it lasts for about four weeks. In the subsong phase a young canary first attempts to sing; the sounds produced are soft in volume and variable. Plastic song is more structured than subsong but is still quite variable. As the bird reaches sexual maturity at seven or eight months, plastic song becomes increasingly stereotyped until it reaches the final phase. The stable-song phase lasts for the duration of the breeding season, at the end of which the bird reverts to plastic song. The return to plastic song is accompanied by a reduction in HVC and RA volume. The transition from plastic song to stable song and back to plastic song, along with the corresponding changes in nuclear volume, are repeated annually in the adult bird. Indeed, the HVC and RA volumes of an adult male canary in late summer are comparable to those of a three-month-old canary, but by the following spring the nuclei have regained their full volume.

also established more contacts, called synapses, with other neurons.

Those anatomical changes suggested that an increase in hormone level or the acquisition of a new behavior (such as singing) or both can act in adulthood to rearrange the distribution of connections between the neurons that control the behavior. Such modifications of existing neural circuits could account for the gross volume changes seen in the RA and could explain why male canaries are able to change their song in adulthood. Males of species such as the white-crowned sparrow and zebra finch, which learn their song before sexual maturity, show no gross changes in adult RA volume.

The developmental and seasonal changes in the HVC and RA volume



ANNUAL SONG VARIABILITY (*green*) in adult male canaries is tied to the blood level (*orange*) of testosterone, a male sex hormone. The song of a canary can be characterized by the number of different sounds (called syllables) it contains. When a bird is in the plastic-song phase, it has not yet settled on a stereotypical song, and it often adds new syllables to its "vocabulary." The addition of new syllables is preceded by a drop in testosterone level. Conversely, few new syllables are incorporated into a canary's vocabulary when its testosterone level is high and it is in the stable-song phase.

seen in male canaries and the marked changes in HVC and RA volume induced by testosterone in adult female canaries suggest that the mastery of learned song is not a subtle process but one that requires conspicuous changes in brain circuitry. Indeed, the seasonal and hormonal changes in the size of song-control nuclei were so extraordinary for an adult vertebrate brain that my co-workers and I were compelled to ask what many neuroscientists considered an unthinkable question: Did the changes really always involve the same set of neurons—those that are present in the brain after sexual maturity?

There is an easy way to determine when a new cell is born. DNA, the substance of which genes are made, is found mainly in a cell's nucleus, and a cell that is about to divide synthesizes new DNA. Hence if one injects an animal with a radioactive form of thymidine, a DNA precursor, the thymidine is sequestered inside the nucleus of cells that are about to undergo division. When such a labeled cell divides, half of the radioactive DNA will be present in the nucleus of each of its two daughter cells, labeling them as well.

Steven A. Goldman and I injected radioactive thymidine into adult male and female canaries every day for several days, stopped the injections and then waited for 30 days. To our surprise, when we then examined the HVC of the birds, we found that as many as 1 percent of the HVC neurons were labeled for each day the birds had received the injections. In another experiment we examined the brains of adult canaries only a day after they had received an injection of radioactive thymidine. In that case we found no labeled neurons in the HVC; we did, however, find many labeled cells in the so-called ventricular zone that overlies the HVC and forms the floor of the lateral ventricle.

These results suggested that the new HVC neurons had been born at the time of the thymidine treatment and that they originated from cells outside the HVC, in the ventricular zone. It appeared that ventricular-zone cells divided into daughter cells that migrated into the HVC, where after a period of between 20 and 30 days they turned into neurons. It so happens that neurons are born in the ventricular zone during development in birds as in all other vertebrates; neurogenesis in adulthood can therefore be viewed as no more than the retention of a developmental trait.

The neurons that develop from ventricular-zone cells look no different from other normal adult-canary neurons. Gail D. Burd and I showed that the new HVC neurons establish synaptic contacts, and John A. Paton and I showed that the new cells generate typical electrical signals when they are stimulated by other neurons. Clearly the new neurons become connected to existing neural circuits as they are added to the adult HVC.

Further studies have shown that new neurons are added constantly to the HVC of adult male and female canaries. Why then does the HVC show no growth from year to year? The obvious answer is that the new neurons replace older ones, which presumably are discarded. Not even the new neurons themselves are exempt from such replacement. My co-workers and I have found very few labeled HVC neurons eight months after the injections of radioactive thymidine, implying that most of the cells do not live much longer than that. Changes in the rates at which new neurons are born and older neurons die could contribute to the seasonal changes in volume of the song-learning nuclei I mentioned above: the number of HVC neurons drops by 38 percent at the end of the breeding season but is fully restored by the following spring.

The next challenge facing investigators is to find out just which neurons are discarded and why. Although neurogenesis and neuron replacement were discovered in a part of the adult canary brain involved in the control of song, it is by no means clear what role the new neurons play in song learning. For instance, the percentage of labeled neurons in the adult HVC per day of treatment with radioactive thymidine is at least as high in nonsinging female as it is in singing male canaries. Labeled neurons have also been found in the HVC of adult male zebra finches that were injected with radioactive thymidine after the end of the critical period, when no song learning takes place.

On the basis of these observations, it seems unlikely that the neurons added to the HVC are just components of the circuits that control the motor skills necessary for song learning. Indeed, physiological experiments suggest that the HVC not only plays an important role in song production but also may have a critical role in song recognition—a perceptual function. Because the HVC may act as a repository of perceptual memories, the addition of new HVC neurons may be necessary for birds to recognize new songs. Just as male songbirds must first acquire a perceptual memory of a model song in order to imitate it, so female birds, as well as male birds past the critical period for song learning, must acquire perceptual memories of new songs in order to recognize the sounds of mates or other birds.

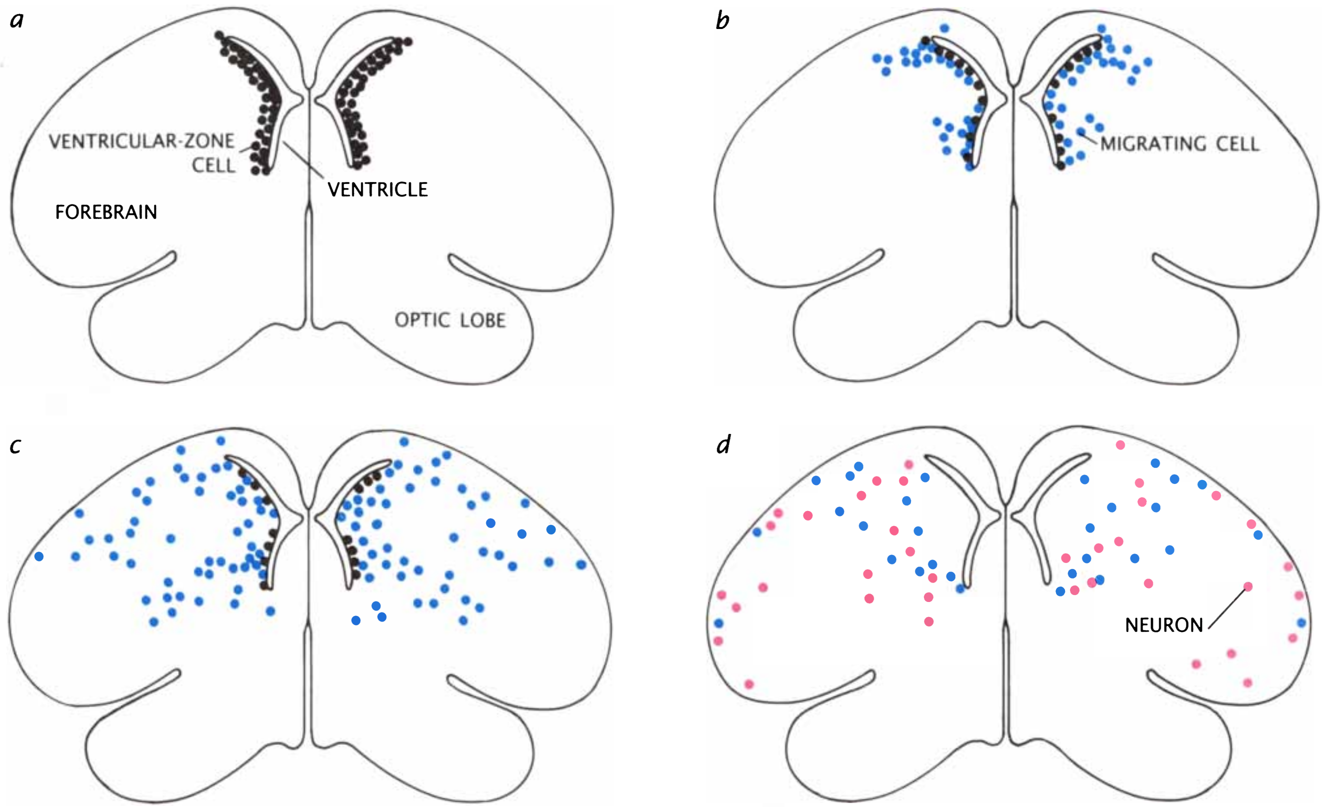
Why must songbirds constantly replenish the neurons in their brain? After all, humans are thought to master motor and perceptual skills with a limited number of irreplaceable neurons. Changes in the synapses of existing neurons could well supply all the neural-circuit flexibility we need for the acquisition of new memories.

Work in my laboratory shows that synaptic changes actually do take place in some parts of the song-control system, such as the RA, where they may relate to motor learning. Yet that flexibility may not be enough for other kinds of learning. Evidence from other laboratories suggests that the kinds of input a brain cell receives can determine which of its genes are expressed, affecting the identity of a cell and the functions it performs. In some cases such genetic changes may underlie a type of irreversible learning. In other words, certain types of neurons involved in song learning may be permanently modified by the memories they hold. Hence the number of available neurons in the avian brain might limit the amount of song learning that can take place. Periodic replacement of HVC neurons may therefore be required in songbirds for them to update perceptual memories of songs.

Neurogenesis in the adult avian brain is not confined to the HVC, although that is the only part of the song-control system to exhibit the phenomenon; it is actually widespread throughout much of the avian forebrain. Interestingly, the forebrain is thought to be the part of the brain most involved with the control of complex learned behaviors.

Although neurogenesis has been reported in some adult mammals, its occurrence there seems to be much more limited and, as I indicated earlier, more controversial. Why is it so prominent in birds? It may have to do with their relatively long life span and airborne lifestyle. A canary weighs as much as a mouse but lives 10 times longer. If a bird had to carry all the brain cells it would need to process and store the information gathered over a lifetime, its brain would have to be substantially larger and heavier.

The widespread occurrence of neu-



MIGRATION AND DIFFERENTIATION of neurons are mapped in this set of cross sections of a canary brain (*see illustration on page 76*) one day (*a*), six days (*b*), 15 days (*c*) and 40 days (*d*) after the cells' birth. The maps are based on data gathered by Arturo Alvarez-Buylla and the author from birds injected with radioactive thymidine, a DNA precursor. The thymidine is taken up by cells that are about to undergo division and is passed on to their daughter cells, which are thereby labeled.

rogenesis in the adult canary forebrain raises another important question. How does a new neuron find its way from its birthplace to its final position in a neural circuit? The answer may be relatively simple for neurons migrating to final sites in the HVC. Since the neurons are generated in the ventricular zone overlying the HVC, they have to traverse distances of no more than half a millimeter before they reach their final positions. Yet many of the new neurons found elsewhere in the avian forebrain are five or six millimeters from the nearest potential birth site. What clues guide a migrating neuron as it travels through the adult brain over a distance that may be 100 times its body length?

Arturo Alvarez-Buylla and I have observed that young neurons migrating from the ventricular wall become elongated and often follow the long fibers of cells known as radial glia, which are very common in the young developing brain of vertebrates as well as in the adult avian forebrain. Although the body of a radial glial cell is lodged in the ventricular wall, its fiber extends into the bulk of the forebrain's gray matter. After several days of migration a young neuron detaches itself from the fiber. We think this occurs when a young neuron approaches the general area where it is meant to assume its adult shape and become part of an existing circuit. Only a third of the migrating cells actually undergo the transformation into a functional, fully differentiated neuron; the rest disappear. A cell's migratory phase lasts for only a few weeks, and it may be that migrating neurons that lose their way during their journey or fail to find a position in a neural circuit in that time simply perish.

Our studies have shown that—contrary to a long-standing doctrine of neurobiology—the brain cells of some adult vertebrates are indeed replaceable. Neurons can be born in adulthood, can travel through the adult vertebrate brain and can take their place in the neural circuits that underlie learning. If the same process could take place in the human brain, it would be invaluable for repairing neural circuits damaged as a result of disease or injury.

Yet there is no evidence that neurogenesis occurs in man or other primates. Perhaps the reason is that humans thrive in the recollection of events long past, and neuronal replacement would disrupt such memories. Nevertheless, it may be possible to induce neurogenesis in adult brains where it does not normally take place. After all, the same genes that orchestrate neurogenesis in the young, developing brain should still be present in the brain cells of an adult. The challenge is to identify the genes and to activate them.

FURTHER READING

NEURONS GENERATED IN THE ADULT BRAIN ARE RECRUITED INTO FUNCTIONAL CIRCUITS. John A. Paton and Fernando Nottebohm in *Science*, Vol. 225, No. 4666, pages 1046–1048; September 7, 1984.

HORMONAL REGULATION OF SYNAPSES AND CELL NUMBER IN THE ADULT CANARY BRAIN AND ITS RELEVANCE TO THEORIES OF LONG-TERM MEMORY STORAGE. Fernando Nottebohm in *Neural Control of Reproductive Function*, edited by Joan M. Lakoski, J. Regino Perez-Polo and David K. Rassin. Alan R. Liss, Inc., 1988.

MIGRATION OF YOUNG NEURONS IN ADULT AVIAN BRAIN. Arturo Alvarez-Buylla and Fernando Nottebohm in *Nature*, Vol. 335, No. 6188, pages 353–354; September 22, 1988.

The Chemical Effects of Ultrasound

Intense ultrasonic waves traveling through liquids generate small cavities that enlarge and implode, creating tremendous heat. These extreme conditions provide an unusual chemical environment

by Kenneth S. Suslick

During the early tests of the first British destroyer in 1894, Sir John I. Thornycroft and Sydney W. Barnaby noticed a severe vibration from the destroyer's propeller. They suggested that large bubbles, or cavities, formed by the spinning propeller and imploded by water pressure, were the source of the vibrations. Thornycroft and Barnaby redesigned the propeller to reduce the vibrations from what came to be known as cavitation, but as the British navy built even faster propulsion systems the phenomenon became an increasingly significant problem. In 1917, therefore, the navy commissioned Lord Rayleigh to study the matter. He confirmed that the vibrations were due to the enormous turbulence, heat and pressure of imploding cavities. Although cavitation continues to be troublesome for nautical engineers, it has provided chemists with a unique environment for high-energy reactions.

Chemistry, after all, is the interaction of energy and matter. Specific energy sources limit the control chemists have over the reactivity of matter. Light interacts with matter on a short time scale at high energies, whereas heat interacts on longer time scales at lower energies. The interaction of sound with matter through the process of cavitation makes available to chemists a range of energies on time scales that are not available from other sources.

Chemists usually induce cavitation not by applying mechanical pressure but rather by generating intense sound waves in a liquid. Such waves create alternating regions of compression and expansion that can form bubbles 100 microns in diameter. The bubbles implode violently in less than a microsecond, heating their contents to 5,500 degrees Celsius—about the temperature of the sun's surface.

The first chemist to recognize the unusual effects of intense sound waves traveling through a liquid, known as sonochemistry, was Alfred L. Loomis in 1927. In spite of his early results, the study of sonochemistry remained exceedingly modest. A renaissance in sonochemistry took place in the 1980's, soon after the advent of inexpensive and reliable generators of high-intensity ultrasound (sound pitched above human hearing at frequencies greater than 16 kilohertz, or 16,000 cycles per second).

Today ultrasound is applied in hospitals for medical imaging, in industry for welding plastics and cleaning materials and even in the home for burglar alarms and vaporizers. These applications, however, do not utilize the chemical effects of ultrasound. Ultrasound can increase the reactivity of metal powders by more than 100,000 times. Ultrasound can drive metal particles together at such high speeds that they melt at the point of collision, and ultrasound can generate microscopic flames in cold liquids.

These chemical effects of ultrasound arise from the physical processes that create, enlarge and implode gaseous and vaporous cavities in a liquid. Ultrasound waves, like all sound waves, consists of cycles of compression and expansion. Compression cycles exert a positive pressure on the liquid, pushing the molecules together; expansion cycles exert a negative pressure, pulling the molecules away from one another.

During the expansion cycle a sound wave of sufficient intensity can generate cavities. A liquid is held together by attractive forces, which determine the tensile strength of a liquid. In order for a cavity to form, a large negative pressure associated with the expansion cycle of the sound wave is needed to overcome the liquid's tensile strength.

The amount of negative pressure needed depends on the type and purity of the liquid. For truly pure liquids, tensile strengths are so great that available ultrasound generators cannot produce enough negative pressure to make cavities. In pure water, for instance, more than 1,000 atmospheres of negative pressure would be required, yet the most powerful ultrasound generators produce only about 50 atmospheres of negative pressure. The tensile strength of liquids is reduced, however, by gas trapped in the crevices of small solid particles. The effect is analogous to the reduction in strength that occurs from cracks in solid materials. When a gas-filled crevice is exposed to a negative-pressure cycle from a sound wave, the reduced pressure makes the gas in the crevice expand until a small bubble is released into solution. Most liquids are sufficiently contaminated by small particles to initiate cavitation. In tap

KENNETH S. SUSLICK is professor of chemistry at the University of Illinois at Urbana-Champaign. He received his B.S. at the California Institute of Technology in 1974 and his Ph.D. from Stanford University in 1978. He holds a Research Career Development Award from the National Institutes of Health and a Sloan Foundation Research Fellowship. In 1986 he was a visiting fellow of Balliol College at the University of Oxford. In addition to sonochemistry, he is interested in the role of metals in biological systems. Suslick also enjoys singing folk music and sculpting bronzes.

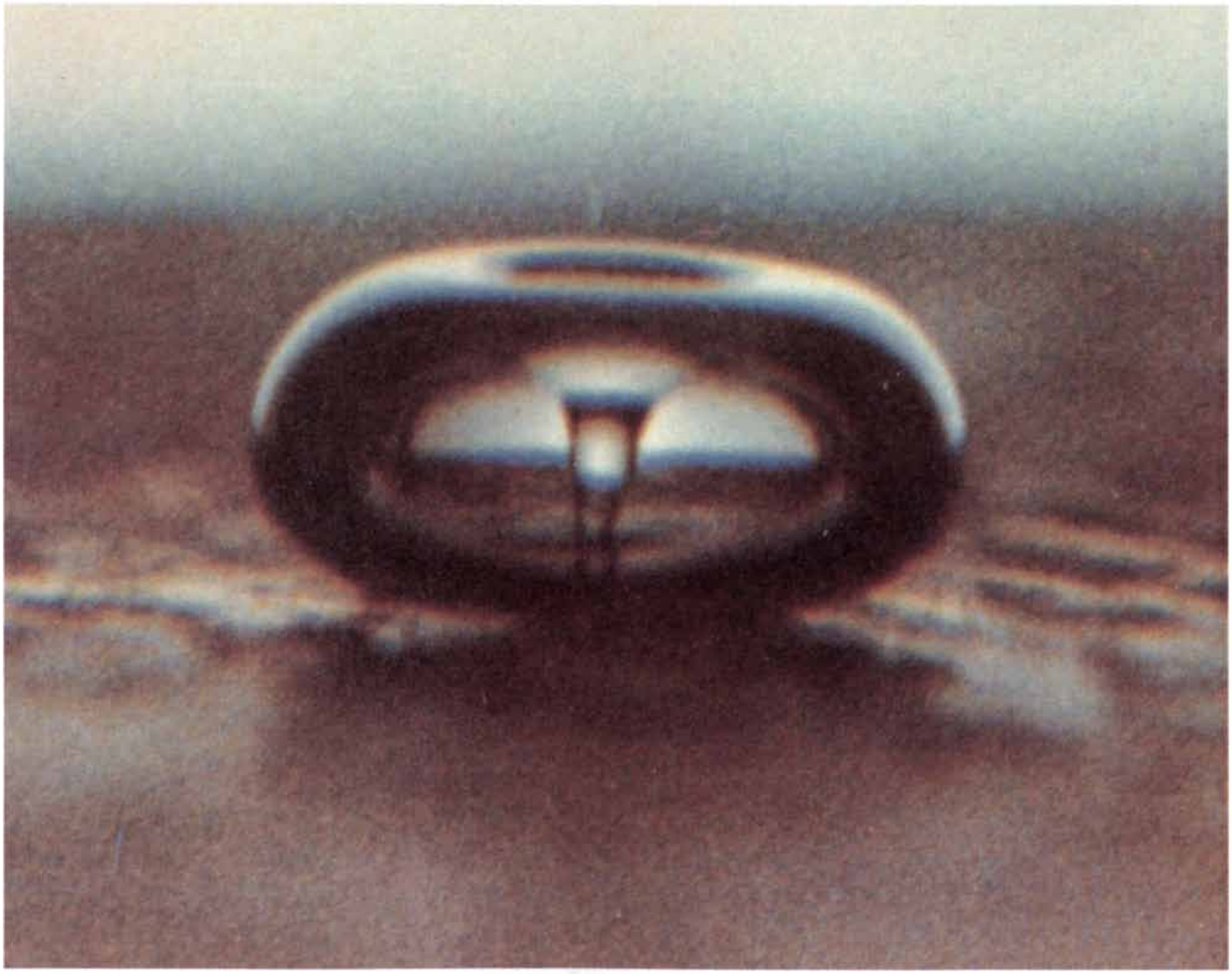
water, negative pressures of only a few atmospheres will form bubbles.

A bubble in a liquid is inherently unstable. If the bubble is large, it will float away and burst at a surface; if it is small, it will redissolve into the liquid. A bubble irradiated with ultrasound, however, continually absorbs energy from alternating compression and expansion cycles of the sound wave. These cause the bubbles to grow and contract, striking a dynamic balance between the vapor inside the bubble and the liquid outside. In some cases the ultrasonic waves will sustain a bubble that simply oscillates in size. In other cases the average size of the bubble will increase.

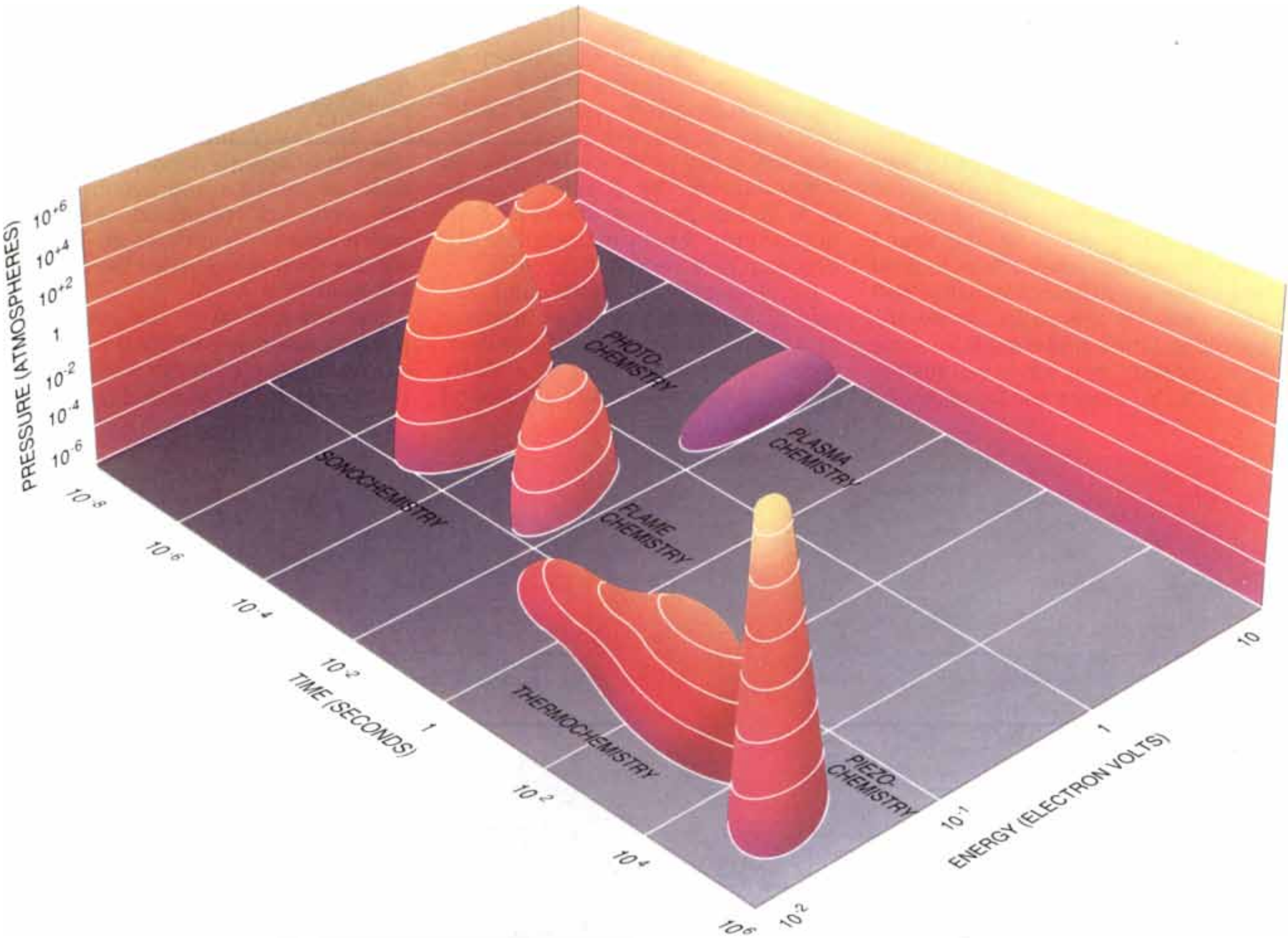
Cavity growth depends on the intensity of sound. High-intensity ultrasound can expand the cavity so rapidly during the negative-pressure cycle that the cavity never has a chance to shrink during the positive-pressure cycle. In this process, therefore, cavities can grow rapidly in the course of a single cycle of sound.

For low-intensity ultrasound the size of the cavity oscillates in phase with the expansion and compression cycles. The surface area of a cavity produced by low-intensity ultrasound is slightly greater during expansion cycles than during compression cycles. Since the amount of gas that diffuses in or out of the cavity depends on the surface area, diffusion into the cavity during expansion cycles will be slightly greater than diffusion out during compression cycles. For each cycle of sound, then, the cavity expands a little more than it shrinks. Over many cycles the cavities will grow slowly.

The growing cavity can eventually reach a critical size where it will most efficiently absorb energy from the ultrasound. The critical size depends on



IMPLODING CAVITY in a liquid irradiated with ultrasound is captured in a high-speed flash photomicrograph (*above*). The implosion heats the gases inside the cavity to 5,500 degrees Celsius. Since this cavity formed near a solid surface, the implosion is asymmetric, expelling a jet of liquid at roughly 400 kilometers per hour. Both the heat and the jet contribute to a unique chemical environment in the liquid. The diameter of the cavity is about 150 microns. The drawing (*left*) shows the stages of cavity implosion and the formation of the jet. The cavity is spherical at first and then shrinks rapidly. The jet develops opposite the solid surface and moves toward it.



SONOCHEMISTRY makes available a range of energies as well as combinations of pressure and duration not available from any other source. The relations among energy, pressure and time are shown for sonochemistry and other chemistry fields.

the frequency of the ultrasound wave. At 20 kilohertz, for example, the critical size is a cavity roughly 170 microns in diameter. At this point the cavity can grow rapidly in the course of a single cycle of sound.

Once a cavity has experienced a very rapid growth caused by either low- or high-intensity ultrasound, it can no longer absorb energy as efficiently from the sound waves. Without this energy input the cavity can no longer sustain itself. The liquid rushes in and the cavity implodes.

The implosion of cavities establishes an unusual environment for chemical reactions. The gases and vapors inside the cavity are compressed, generating intense heat that raises the temperature of the liquid immediately surrounding the cavity and creates a local hot spot. Even though the temperature of this region is extraordinarily high, the region itself is so small that the heat dissipates quickly. My co-workers and I at the University of Illinois at Urbana-Champaign estimate that the heating and cooling rates during cavitation are more than a billion degrees C per second! This is similar to the cooling that occurs if molten metal is splattered onto a surface cooled near absolute zero. At any given time, therefore, the bulk of the liquid remains at the ambient temperature.

The exact temperatures and pressures generated during cavity implosion are difficult both to calculate theoretically and to determine experimentally. Yet these quantities are fundamental to describing the potential of sonochemistry. Theoretical models have been proposed that approximate the dynamics of cavity implosion at various levels of accuracy. All have difficulty accurately describing cavity dynamics during the last stages of implosion. The most sophisticated models give temperatures of thousands of degrees Celsius, pressures of from hundreds to thousands of atmospheres and heating times of less than a microsecond.

The temperature of the imploding cavity cannot be measured with a physical thermometer because the heat is dissipated too quickly to be registered. Recently my collaborators and I found an alternative that enables one to check theoretical results experimentally. One way to gauge temperature is to observe the rate at which familiar chemical reactions take place. More precisely, the temperature is related to the negative inverse logarithm of the rate of the reaction. If the rates of several different reactions are measured in an ultrasound environment, the temperature from cavity implosion can be calculated.

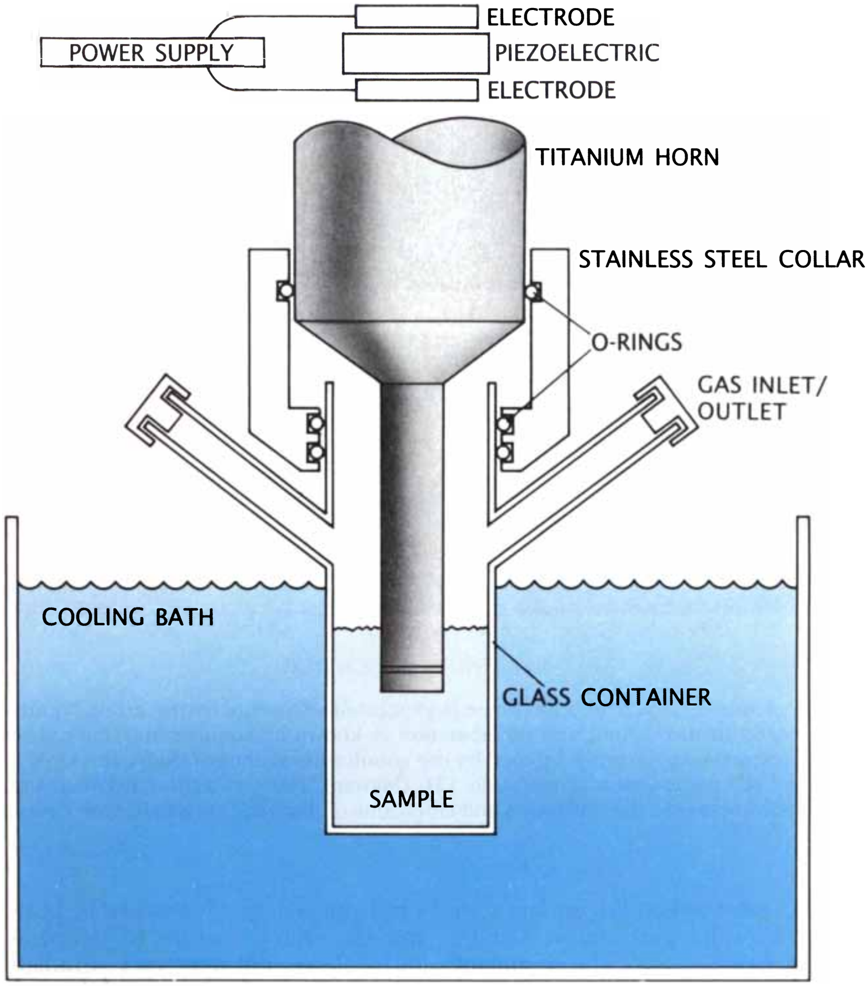
In the process of determining the relative rates of a series of sonochemical reactions, David A. Hammerton in our laboratory discovered two distinct temperature regions associated with cavity implosion. The original gaseous contents of the cavity reached temperatures of about 5,500 degrees C, whereas the liquid immediately surrounding the cavity reached 2,100 degrees. The temperature of the flame from an acetylene torch, in comparison, is about 2,400 degrees.

Although the pressures attained during cavity implosion are harder to determine experimentally than temperature, the two quantities are correlated. One can therefore estimate the peak pressure to be 500 atmospheres, which is half the pressure at the deepest region of the ocean, the Mariana Trench.

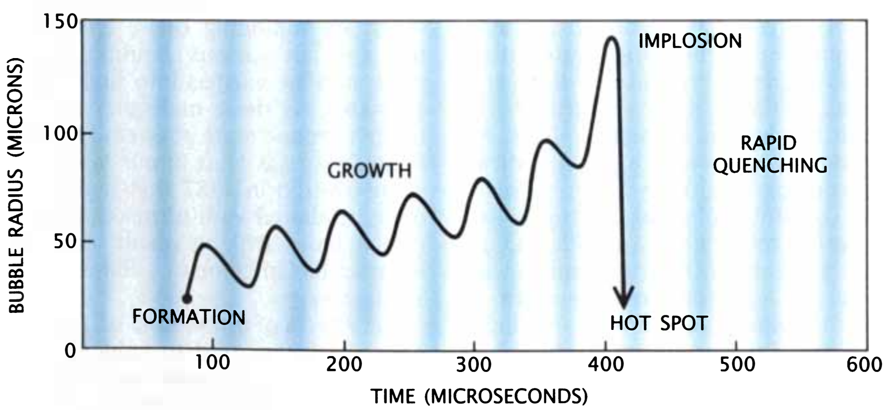
Even though the local temperature and pressure conditions created by cavity implosion are extreme, chemists have good control over sonochemical reactions. The intensity of cavity implosion, and hence the nature of the reaction, can easily be altered by such factors as acoustic frequency, acoustic intensity, ambient temperature, static pressure, choice of liquid and choice of ambient gas. The way these factors alter sonochemistry often defies basic intuitions about chemistry. Let me cite three examples.

First, unlike virtually all chemical reactions, most sonochemical reactions decrease in rate with increasing ambient temperature—that is, the temperature outside the cavity. The higher the ambient temperature is, the more vapor there will be inside the cavity. The extra vapor cushions the implosion of the cavity and lowers the temperature of implosion. Therefore sonochemical reactions proceed more slowly as ambient temperature increases. Second, unlike chemical events driven by light, sonochemical reactions do not depend greatly on frequency. The major effect of frequency is to change the critical size of a cavity before implosion, which does not change the cavitation process significantly. Third, unlike many chemical reactions in solution, the ambient gas that is dissolved in the liquid is quite important. If xenon fills a cavity, the peak temperature reached during cavity implosion will be high because xenon conducts heat poorly and retains the heat of the collapsing cavity. Helium, on the other hand, conducts heat so well that it can virtually shut down sonochemical reactions.

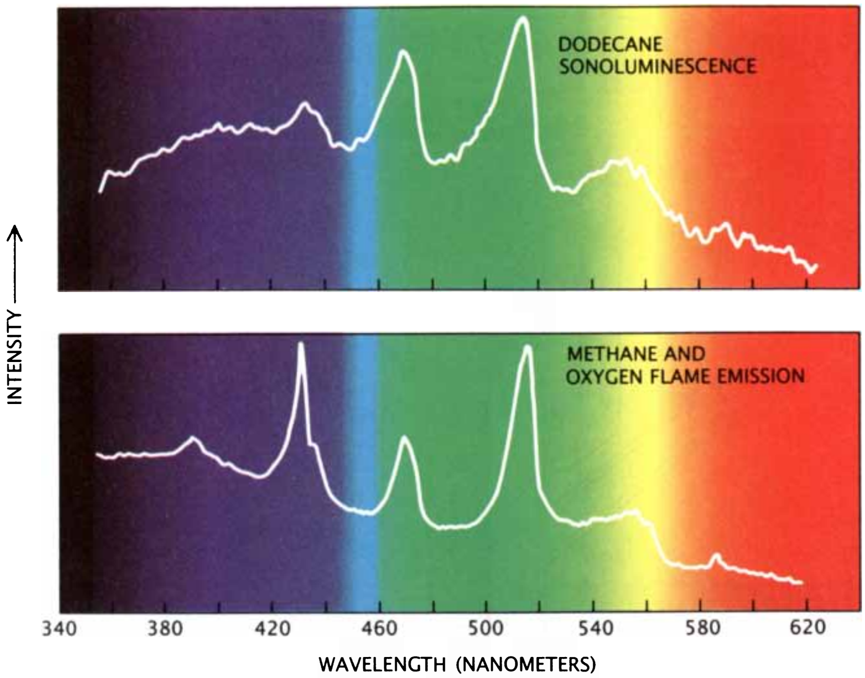
The dynamics of cavity growth and implosion are strongly dependent on local conditions, including the form of the materials: whether they are



ULTRASONIC IMMERSION HORN is the intensest generator of ultrasound in general use. Ultrasound is produced in liquids by means of piezoelectric or magnetostrictive materials: materials that expand or contract when they are placed in electromagnetic fields. Exposing such materials to a field alternating at an ultrasonic frequency produces ultrasound. In the ultrasonic horn there is a piezoelectric ceramic attached to a tapered titanium rod, which serves to amplify the sound. The horn vibrates at a fixed acoustic frequency, typically 20 kilohertz, but the intensity is variable. Since power outputs are quite high, the reaction solution must be cooled.



BUBBLE GROWTH AND IMPLOSION in a liquid irradiated with ultrasound is the physical phenomenon responsible for most sonochemistry. Intense ultrasound waves generate large alternating stresses within a liquid by creating regions of positive pressure (*dark color*) and negative pressure (*light color*). A cavity can form and grow during the episodes of negative pressure. When the cavity attains a critical size, the cavity implodes, generating intense heat and tremendous pressure.



LIGHT resembling that from a gas flame is generated when cold hydrocarbon liquids are exposed to ultrasound; the phenomenon is known as sonoluminescence. The graphs show the spectrum produced by the sonoluminescence of dodecane, $C_{12}H_{24}$ (*top*), and the combustion of methane, CH_4 (*bottom*). The similarities between the spectrums are due to the formation and emissions of diatomic carbon in both cases.

liquids, extended solid surfaces in liquids or solid particles in liquids. In each case, because the dominant physical effects from cavity implosion differ, the chemistry changes as well.

The sonochemistry of liquids depends mainly on physical effects of the quick heating and cooling caused by cavity implosion. For instance, when Peter Riesz and his colleagues at the National Cancer Institute irradiated water with ultrasound, they proved that the heat from cavity implosion decomposes water (H_2O) into extremely reactive hydrogen atoms (H^+) and hydroxl radicals (OH^-). During the quick cooling phase, hydrogen atoms and hydroxl radicals recombine to form hydrogen peroxide (H_2O_2) and molecular hydrogen (H_2). If other compounds are added to water irradiated with ultrasound, a wide range of secondary reactions can occur. Organic compounds are highly degraded in this environment, and inorganic compounds can be oxidized or reduced.

Other organic liquids also yield interesting reactions when they are irradiated with ultrasound. For example, alkanes, major components of crude oil, can be "cracked" into smaller, desirable fragments, such as gasoline. Crude oil is normally cracked by heating the entire mixture to temperatures above 500 degrees C. Irradiating alkanes with ultrasound, however, makes cracking possible at room temperature and produces acetylene, which cannot be produced through simple heating.

Perhaps the most unusual chemical phenomenon associated with ultrasound is its ability to produce microscopic flames in cold liquids by a process known as sonoluminescence. When an imploding cavity creates a hot spot in various liquids, molecules may be excited into high-energy states. As these molecules return to their ground state, they emit visible light. Edward B. Flint in our laboratory discovered in 1987 that hydrocarbons irradiated with ultrasound provide a most striking result: emitted light similar in color to a flame from a gas stove.

The effects of ultrasound on liquids have also been used to enhance the chemistry of compounds in solution. Compounds that contain metal-carbon bonds, called organometallics, are particularly illustrative. This diverse class of chemicals is important in the formation of plastics, in the production of microelectronics and in the synthesis of pharmaceuticals, herbicides and pesticides. In 1981 Paul F. Schubert and I first investigated the effects of ultrasound on organometallics, in particular iron pentacarbonyl, or $Fe(CO)_5$. The results, when compared with the effects of light and heat on $Fe(CO)_5$, underscore the distinctive chemistry that ultrasound can induce [*see illustration on opposite page*]. When $Fe(CO)_5$ is exposed to heat, it decomposes into carbon monoxide (CO) and a fine iron powder, which ignites spontaneously in air. When $Fe(CO)_5$ is exposed to ultraviolet light, it first breaks down into $Fe(CO)_4$ and free CO fragments. $Fe(CO)_4$ can then recombine to form $Fe_2(CO)_9$. Cavity implosion creates different results. It delivers enough heat to dissociate several CO molecules but cools quickly enough to quench the reaction before decomposition is complete. Thus when $Fe(CO)_5$ is exposed to ultrasound, it yields the unusual cluster compound $Fe_3(CO)_{12}$.

The sonochemistry of two immiscible liquids (such as oil and water) stems from the ability of ultrasound to emulsify liquids so that microscopic droplets of one liquid are suspended in the other. Ultrasonic compression and expansion stress liquid surfaces, overcoming the cohesive forces that hold a large droplet together. The droplet bursts into smaller ones, and eventually the liquids are emulsified.

Emulsification can accelerate chemical reactions between immiscible liquids by greatly increasing their surface contact. A large contact area enhances crossover of molecules from one liquid to the other, an effect that can make some reactions proceed quickly. Emulsifying mercury with various liquids has particularly interesting chemistry as delineated by the investigations of Albert J. Fry of Wesleyan University. He developed the reactions of mercury with a variety of organobromide compounds as an intermediate in the formation of new carbon-carbon bonds. Such reactions are critical in the synthesis of complex organic compounds.

The sonochemistry of solid surfaces in liquids depends on a change in the dynamics of cavity implosion. When cavitation occurs in a liquid near an extended solid surface, the cavity implosion differs substantially from the symmetrical, spherical implosion observed in liquid-only systems. The presence of the surface distorts the pressure from the ultrasound field so that a cavity implosion near a surface is markedly asymmetric. This generates a jet of liquid directed at the surface that moves at speeds of roughly 400 kilometers per hour. The jet, as well as the shock waves from cavity implosion, erode solid surfaces, remove nonreactive coatings and fragment brittle powders. Reactions are further facilitated by high temperatures and pressures associated with cavity implosion near surfaces. These processes all enhance the chemical reactivity of solid surfaces, which is important in the synthesis of drugs, specialty chemicals and polymers.

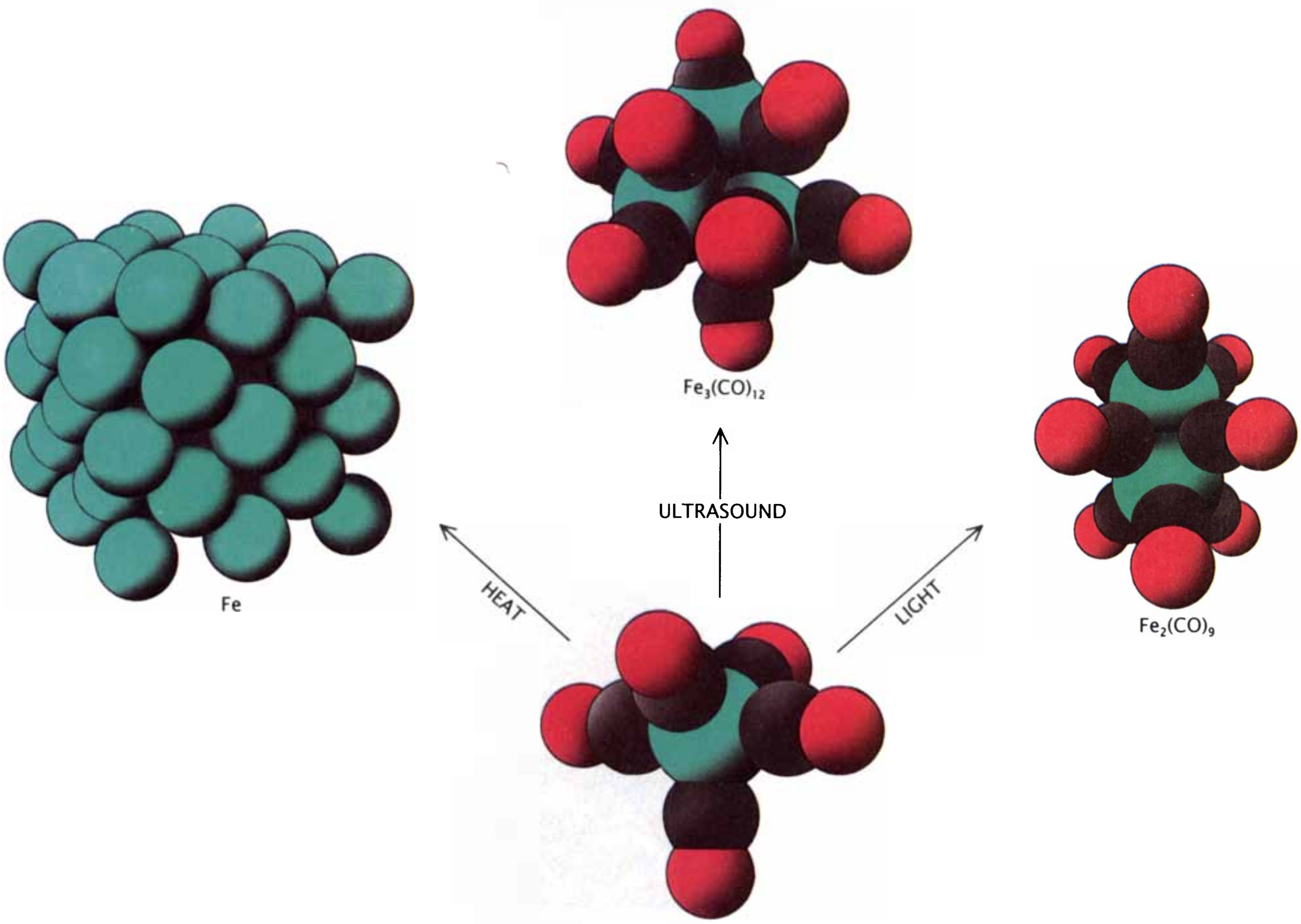
The sonochemistry of solid surfaces in liquids is best exemplified by reactions of active metals, such as lithium, magnesium, zinc and aluminum. Ultrasonic irradiation of reaction mixtures constituting these metals provides better control at lower temperatures and produces relatively higher yields. Pierre Renaud of the University of Paris first examined such reactions. More recently Jean-Louis Luche of the University of Grenoble and Philip Boudjouk of North Dakota State University have popularized the use of an ultrasonic cleaning bath to accelerate the reactions of active metals.

The chemistry of these metals is very difficult to control. Traces of water, oxygen or nitrogen can react at the surface to form protective coatings. Increasing the reactivity of the protected surface by direct heating, however, can result in undesirable explosions. Ultrasound can keep the surface clean and allows the reaction to proceed evenly at reduced ambient temperatures. Excellent yields and improved reliability can be achieved for many reactive metals in large-scale industrial applications.

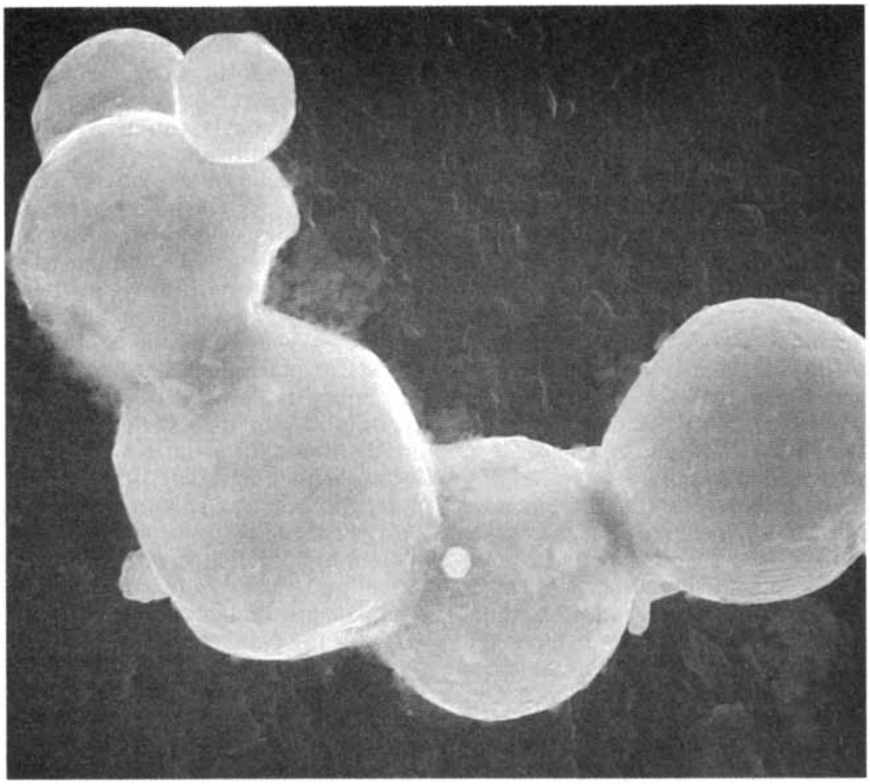
The extreme conditions generated by cavitation near surfaces can also be utilized to induce reactivity in "unreactive" metals. Robert E. Johnson in our laboratory, for instance, examined reactions between carbon monoxide and molybdenum and tantalum, as well as other comparable metals. Conventional techniques require pressures of from 100 to 300 atmospheres and temperatures of from 200 to 300 degrees C. to form metal carbonyls. Using ultrasound, however, formation of metal carbonyls can proceed at room temperature and pressure.

The implosion of a cavity, in addition to all the effects described so far, sends shock waves through the liquid. The sonochemistry of solid particles in liquids depends heavily on these shock waves; they drive small particles of a powder into one another at speeds of more than 500 kilometers per hour. My co-workers and I have recently shown that such collisions are so intense in metal powders that localized melting takes place at the point of impact. This melting improves the metal's reactivity, because it removes metallic-oxide coatings. (Such protective oxide coatings are found on most metals and are responsible for the patina on copper gutters and bronze sculpture.)

Since ultrasound improves the reactivity of metal powders, it also makes



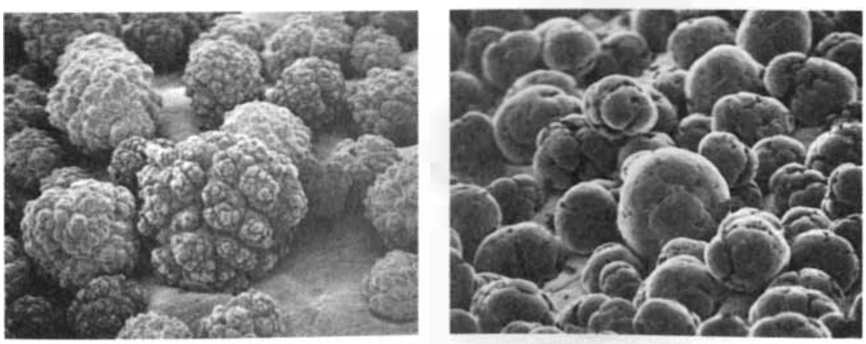
CHEMISTRY OF ULTRASOUND can differ greatly from the chemistries of light and heat. The reactions of iron pentacarbonyl $Fe(CO)_5$, an iron atom (*green*) bonded to five carbon monoxide units (*gray and red*), exemplify the differences. Heat decomposes iron pentacarbonyl into pure iron and carbon monoxide. Light yields two iron atoms bonded to nine carbon monoxide units. Ultrasound produces the cluster compound consisting of three iron atoms bonded to 12 carbon monoxide units.



INTERPARTICLE COLLISIONS driven by shock waves from cavity implosion cause particles to agglomerate. The collisions between particles of zinc (here magnified 13,000 diameters) were so violent that the zinc particles melted on impact.

them better catalysts. Many reactions require a catalyst in order to proceed at useful or even appreciable rates. Catalysts are not consumed by the reaction but instead speed the reaction of other substances.

The effects of ultrasound on particle morphology, surface composition and catalyst reactivity have been investigated by Dominick J. Casadonte and Stephen J. Doktycz in our laboratory. They have discovered that catalysts such as nickel, copper and zinc powders irradiated with ultrasound show dramatic changes in surface morphology. Individual surfaces are smoothed and particles are consolidated into extended aggregates. An experiment to determine the surface composition of nickel revealed that its oxide coating was removed, greatly improving the reactivity of the nickel powder. Ultrasonic irradiation increased the effectiveness of nickel powder as a catalyst more than 100,000 times. The nickel powder is as reactive as some special catalysts currently in use, yet it is nonflammable and less expensive.



METAL SURFACES can be physically altered by ultrasound. Before this sample of nickel powder was irradiated with ultrasound (*left*) it had an unreactive crystalline coating on its surface. After irradiation (*right*) the coating was gone, exposing the nickel and boosting its reactivity. Without the coating the nickel powder becomes an excellent catalyst for chemical reactions. The magnification is about 100 diameters.

Ultrasound is a useful tool in nearly every case where a liquid and a solid must react. Furthermore, since ultrasound can radiate through large volumes of liquid, it is well suited for industrial applications. For these reasons future applications of ultrasound in chemical reactions will be diverse. In the synthesis of pharmaceuticals, ultrasound will improve chemical yields over conventional methods.

The greatest advances in sonochemistry, however, may be in the production of new materials that have unusual properties. The extraordinary temperatures and pressures reached during cavitation, for example, may lead to the synthesis of refractory materials (such as carborundum, tungsten carbide and even diamond). Refractory solids have high temperature stability and enormous structural strength. They are important as industrial abrasives and hardened tool bits.

The extremely rapid cooling that follows cavity implosion may be employed to create metallic glasses. Such amorphous metals have outstanding corrosion resistance and unusually high strength.

Although chemical applications of ultrasound are still in the early stages of development, the next few years promise rapid progress in sonochemistry. The use of ultrasound in laboratory reactions is becoming commonplace, and the extension of the technology to industrial-scale reactions is likely to follow. Underlying these developing technologies are the recent advances in our understanding of the nature of cavitation and the chemical effects of ultrasound.

FURTHER READING

ULTRASOUND IN SYNTHESIS. Kenneth S. Suslick in *Modern Synthetic Methods,* Vol. 4, pages 1-60; 1986.

SYNTHESIS WITH ULTRASONIC WAVES. Philip Boudjouk in *Journal of Chemical Education,* Vol. 63, No. 5, pages 427-429; May, 1986.

SONOLUMINESCENCE FROM NON-AQUEOUS LIQUIDS. Kenneth S. Suslick and Edward B. Flint in *Nature,* Vol. 330, No. 6148, pages 553-555; December 10, 1987.

ULTRASOUND: ITS CHEMICAL, PHYSICAL AND BIOLOGICAL EFFECTS. Edited by Kenneth S. Suslick. VCH Publishers, Inc., 1988.

Teeth and Prehistory in Asia

Minute differences in tooth structure enable the author to reconstruct the great prehistoric migrations that peopled the New World as well as east Asia and the Pacific Basin

by Christy G. Turner II

When one thinks of evidence for the great migrations of human prehistory, what usually comes to mind is artifacts: a stone knife, a bit of house timber, a piece of pottery. Yet other, humbler, kinds of evidence can be equally informative. Teeth, for example. The precise form of human teeth results from patterns of genetic inheritance that remain stable from generation to generation within a given population. Therefore differences among groups can be used to decipher how they are related, much in the manner of a family tree. For three decades my work has focused on using dental evidence to understand the prehistoric migrations that peopled Asia, the Pacific Basin and the Americas; by now a reasonably clear picture has emerged.

This picture is based on a twofold concept I call Sundadonty and Sinodonty. Each of these represents a distinct group of dental traits, and they correspond to the two large branches of a population system. Sundadonts now people southeast Asia, Indonesia and Polynesia; Sinodonts populate China, Japan and Siberia as well as all of the New World. The center of this great web of humanity appears to have been southeast Asia. It was from there that the people who would become Sinodonts spread to China, then to Siberia and eventually across the Bering land bridge to North America. It is only because of the stability of dental patterns that we can trace this fine web all the way from the Paleo-Indians of Chile to the web's origins in southeast Asia.

CHRISTY G. TURNER II is professor of anthropology at Arizona State University. Dental anthropology has been the focus of his work since he completed his doctoral research in 1967 on the dentition of the American Arctic peoples. He is a past fellow of the Center for Advanced Studies in the Behavioral Sciences and Arizona State Research Professor. His data collecting (in 16 countries) has been assisted by many organizations, including the National Geographic Society, the National Science Foundation, the National Academy of Sciences of the U.S., the Academy of Sciences of the U.S.S.R. and IREX.

Following the European explorations of the Pacific, east Asia and the Americas from about 1500 to 1750, many ideas were proposed to explain the origins and relations of the many peoples who inhabit that vast nameless region. Some of these ideas were reasonable, others absurd. Years ago Polynesians were thought to have originated in Egypt. The Ainu of northern Japan have long been thought to be some kind of wayward Caucasoids. Negritos of the Philippine and Andaman islands are still sometimes said to have African origins. Many places of origin have been offered for American Indians and Eskimos, including now sunken islands in the Pacific, Antarctica, the Mediterranean and even Europe.

Part of the disagreement and the "soft," speculative nature of these ideas stem from the great diversity of the populations in this region, which encompasses about half of the earth's surface. Within it are tropical islands, lush rain forests, hot dusty steppes, icy arctic tundra, boreal forests, deserts, ocean coasts, mountains and prairies. And in these dissimilar environments an equally bewildering array of cultures has taken root among the Pacific, American and Asian populations.

Yet not all the difficulties lie in the complexity of the human populations themselves. Some stem from the kind of data that have been selected in the effort to puzzle out the relations among groups. Some scholars have relied on acquired traits, such as styles of toolmaking technology, to compose the family tree. Acquired traits, however, are subject to much more environmental influence than those that are genetically inherited. Isolated cultural elements such as fishhooks, houses or hats may converge toward striking stylistic similarity in the absence of a common origin.

Even if genetically inherited traits are chosen, the problem is not automatically resolved. Traits (eye color, for instance) that depend on the inheritance of a single gene can change radically in their frequency in the course of a migratory move, particularly if the trait has a low frequency in the original population. If a trait has a frequency of, say, 10 percent in the source population, and the group that leaves to colonize a new territory is small, then it is quite likely that the trait will be altogether absent among the colonizers. Thus, for example, although the B blood type is common in northeast Asia, it is almost entirely absent among the Native American population—which is descended from prehistoric inhabitants of northeast Asia.

Therefore the best variables for reconstructing population history are those that depend on not one but many genes and vary minimally in response to environmental influences. For practical reasons it is also helpful if the traits are easy and inexpensive to examine and can be compared directly between living and ancient populations. Certain aspects of the skeleton, in particular the teeth, meet these requirements. As a result "dental anthropology" has been a well-established discipline for almost half a century, dating from the pioneering studies of Ales Hrdlicka of the Smithsonian Institution and others.

What are the dental characteristics that enable anthropologists to distinguish among human populations? Human teeth have five basic features. First is the crown, armored with hard enamel for cutting and milling food. Second are the long roots that anchor each tooth in the supporting bone. Third is the division of the teeth into types: incisors, canines, premolars and molars. Fourth is the number of teeth—32—each of which (except for the molars) is preceded by one of the 22 deciduous teeth of childhood.

Now, all these general features are found in all modern human beings, demonstrating a strikingly stable polygenetic inheritance. What varies much more within and between major human divisions are the frequencies of the secondary traits. These secondary traits include the number of cusps (rounded bumps on the biting surface) on the molars, the number of roots, various tiny ridges and grooves in the enamel and other small anatomical features. Although the frequencies of these secondary traits differ considerably between groups, they are relatively stable over time within a group—even quite a small one—as has been shown in numerous studies by various workers.

I have employed these secondary traits to develop a concept that provides a unifying evolutionary hypothesis about the biocultural and linguistic variation of the peoples of the Pacific Basin and Rim. Actually this concept occurred to me in a single insight about 15 years ago while I was making dental observations with Kazuro Hanihara of the University of Tokyo. It suddenly seemed that the great mosaic of Pacific peoples actually has only two major pieces: Sundadonty and Sinodonty. Among the Sundadonts are the aboriginal peoples of southeast Asia, including the Thais, Andaman Islanders and Burmese, the early Cambodians and Laotians, Malaysians, Philippians, the Taiwan and Borneo aborigines and other Indonesian peoples.

Sinodonty, on the other hand, is a slightly different Asian dental pattern found among the more northerly and arctic east Asian populations: Chinese, Mongols, Buryats, modern Japanese, eastern Siberians and others, including all the original inhabitants of the Americas. The boundary between the two main patterns seems to lie in southern China. Since arriving at this twofold conception, I have observed

PEG MOLAR, shown at the left, is common among Aleuts and Eskimos, typically as the upper third molar. It is one of the characteristics that has enabled the author to divide all the native peoples of east Asia and the New World into two large groups: Sundadonts and Sinodonts. Aleuts and Eskimos (along with all other Native American peoples) are Sinodonts.

many more thousands of teeth and done a variety of statistical analyses; the concept continues to hold up.

Of the 30 or so well-known and commonly studied secondary traits of human teeth, the frequencies of eight have the most statistical significance in distinguishing Sundadonty and Sinodonty. In the upper jaw there are incisor shoveling and double-shoveling, first-premolar root number, first-molar enamel extension and third-molar reduction, and in the lower jaw there are first-molar deflecting wrinkle, first-molar root number and second-molar cusp number. (Since these traits are most readily understood through pictures, the most important of them are displayed in the illustrations accompanying this article.)

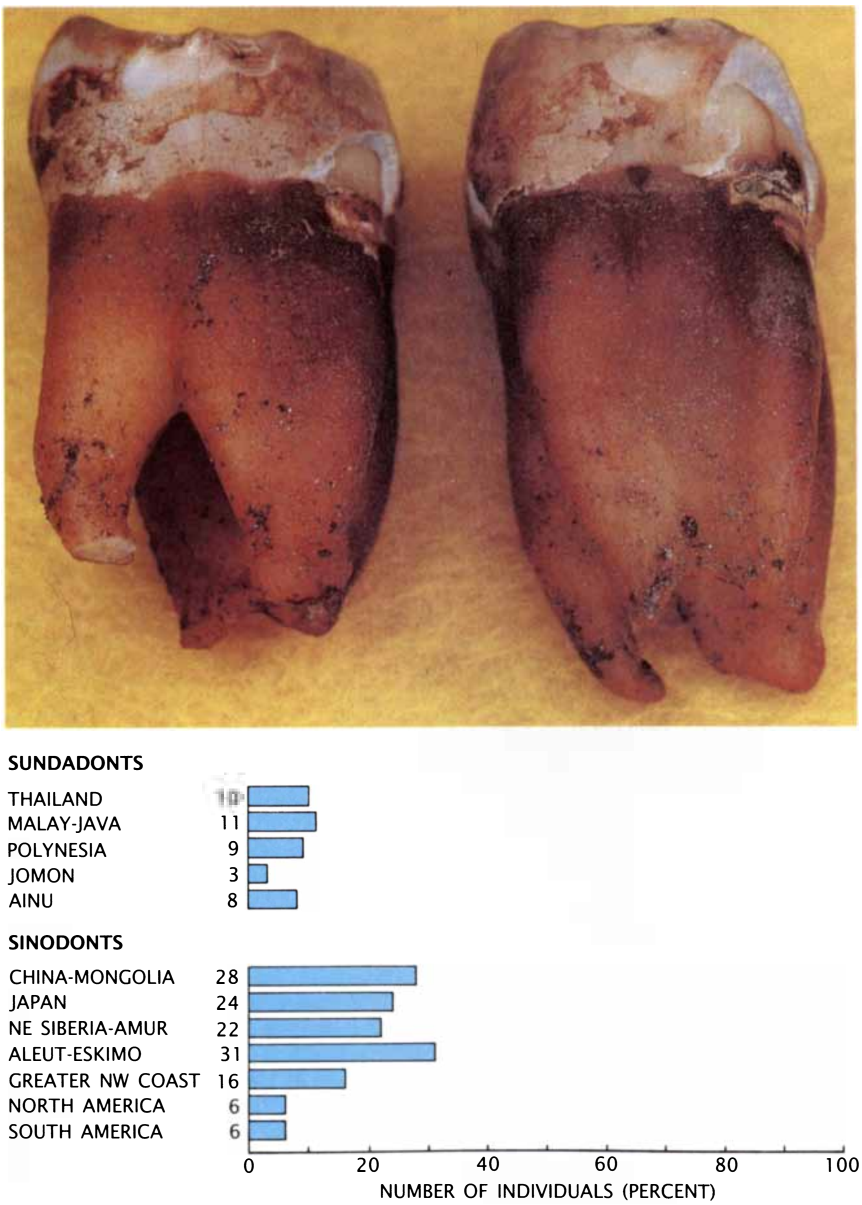
All the populations that make up the Sundadonts and Sinodonts are generally recognized as members of the Mongoloid geographic race. Yet it is evident that there are variations among these groups that have great utility for biohistorical interpretation. These differences appear as variations in the frequency of particular secondary dental traits. Take upper-incisor shoveling, for example, which is the presence of additional marginal ridges in the crown, giving the incisor a "shovel" appearance [*see illustration on opposite page*].

When the groups are compared, some notable differences in frequency of shoveling are seen. Southeast Asian Thais and Malay-Javanese have significantly lower frequencies of shoveling than northeast Asians or Native Americans do. The Polynesians, the Jomonese (ancient inhabitants of Japan) and the Ainu (modern inhabitants of some parts of northern Japan whose extraction differs from that of the majority of modern Japanese) also have low shoveling frequencies, whereas the dominant Japanese population has a high frequency.

Shoveling frequency thus defines two population groups—Sundadonts and Sinodonts. Repeated statistical analyses of these and other population samples always yields the same results: the Polynesians, the Jomonese and the Ainu cluster with the southeast Asians as Sundadonts, whereas the Japanese consistently join the northeast Asians and the Native American population as Sinodonts.

These dentally defined relations among groups tell an intriguing evolutionary story. The tale begins sometime before 50,000 years ago in southeast Asia. The earliest anatomically modern humans in that region probably came from Africa, but they may have evolved independently. Whatever their exact origins, they shared a generalized dental pattern with other early modern humans. The oldest known human skeletal remains representing this pattern in southeast Asia are the Tabon find in the Philippines, which is 20,000 years old, and the Niah Cave skull from Sarawak, 40,000 years old; both specimens are older than any known modern-human find in northeast Asia.

Out of the generalized early dental pattern shown by the Tabon and Niah Cave specimens there developed Sundadonty. Sundadonty is closely related to the early, generalized pattern, retaining its crown and root characteristics, with some crown simplification. The centerpiece of the region where Sundadonty developed is the Sunda Shelf, a continental plain that once bridged island and mainland southeast Asia. At the time the Sundadont pattern developed, sea level was 100 meters lower than it is today. Beginning near the end of late Pleistocene times the sea rose and by 12,000 years ago the Sunda Shelf was largely flooded, leaving the Indonesian archipelago, Japan, island southeast Asia and other areas above water.



THREE ROOTS are found on a lower first molar when an extra root develops along with the usual two; a three-rooted tooth is shown at the left. As the chart indicates, such three-rooted molars are commoner among Sinodonts than among Sundadonts.

When the sea rose, the Jomonese people, who were Sundadonts, were isolated in Japan, whence they had migrated by way of the coast of the now flooded east Asian continental shelf. Hence one can conclude that Sundadonty is at least 12,000 years old. In fact, however, the pattern is likely to be considerably older. The dental pattern of the 17,000-year-old Minatogawa skeletons found on the island of Okinawa and studied by Hanihara and Hisashi Suzuki of the University of Tokyo is almost certainly Sundadont; the only reason for doubt is the small size of the sample. The modern-day Ainu closely resemble the Jomonese, and the sequence of descent was likely to have been from southeast Asians to Minatogawa to Jomon to living Ainu.

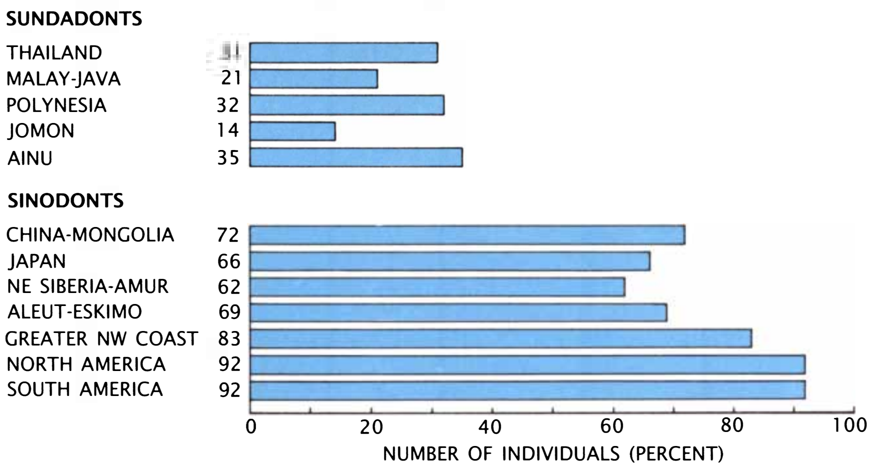
A bit more light is thrown on the origins of Sundadonty by considering the natives of Australia. The ancestors of Australian aboriginals must have been in southeast Asia, if only briefly, before expanding through Indonesia into Australia. The earliest people known to reach Australia rafted there some 30,000 years ago. These earliest arrivals do not have the Sundadont pattern; they have the more generalized early modern pattern. Therefore it seems likely that Sundadonty developed between 30,000 and 17,000 years ago (when it is observed in the Minatogawans of Okinawa).

How did the Sundadont pattern spread and change after its inception? While some Sundadonts (such as the Jomon) were migrating north along the coast of the continent, others were moving north in the interior. The interior expansion led to the formation of the northeast Asian populations among whom the Sinodont pattern evolved. Sinodonty exhibits specialization and intensification of several features of the roots and crowns of the teeth. These changes could have been microevolutionary consequences of adaptation to life in the colder, more stressful northern regions. It is more likely, however, that they were simply random genetic changes that took place in small, isolated hunting-and-fishing groups of the late Pleistocene in northeast Asia.

People possessing the Sundadont pattern expanded into China and Mongolia about 20,000 years ago, and that is where I suspect Sinodonty rapidly developed. The new pattern is evident in craniums found at Zhoukoudian Upper Cave in northern China, a site that is more than 11,000 years old, and Sinodonty must have been evolving for some time before that.

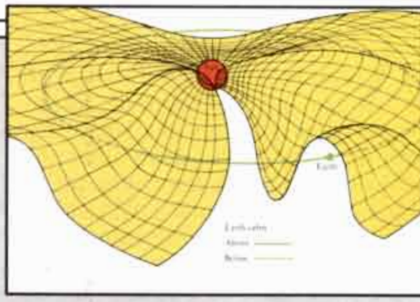
About 2,000 years ago the Sinodont pattern was carried from the mainland of Asia to Japan. The initial migration of the people who were to become the modern Japanese took place in about 200 B.C., as evidenced by the appearance of wet-rice farming, metallurgy and new settlement patterns in a new culture called the Yayoi. Yayoi teeth closely resemble those of the modern Japanese, but they differ markedly from the Jomonese (the Sundadonts who had migrated along the coast to Japan 10 millenniums before) or the Ainu, their probable descendants. Recent linguistic research by Paul K. Benedict of Arizona State University suggests that the Japanese language originated in southern China. His work is complemented by mine: I find that Japanese teeth most closely resemble those of the southern China people in and around Hong Kong.

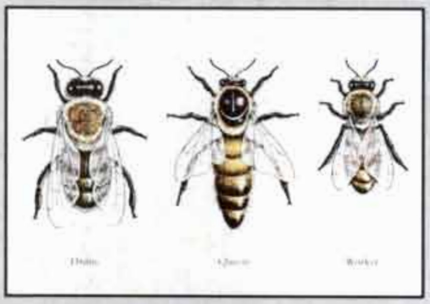
Long after the Sundadont expansion that ultimately yielded Sinodonty, there was another migration of the Sundadont population. This expansion, only centuries before the beginning of the Christian Era, carried families and relatives of seafaring agriculturists across most of the Pacific. We

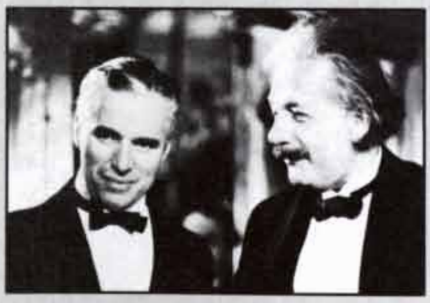


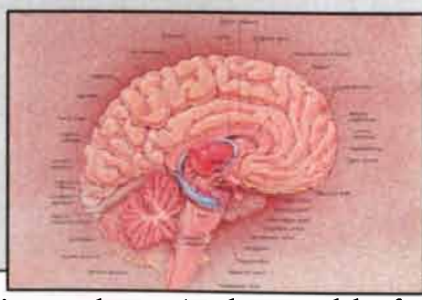
SHOVELING is a condition of the inner surface of the upper incisors created by the presence of extra ridges that give the crown a "shovel" appearance. As the chart shows, shoveling much is more frequent among Sinodonts than among Sundadonts.

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

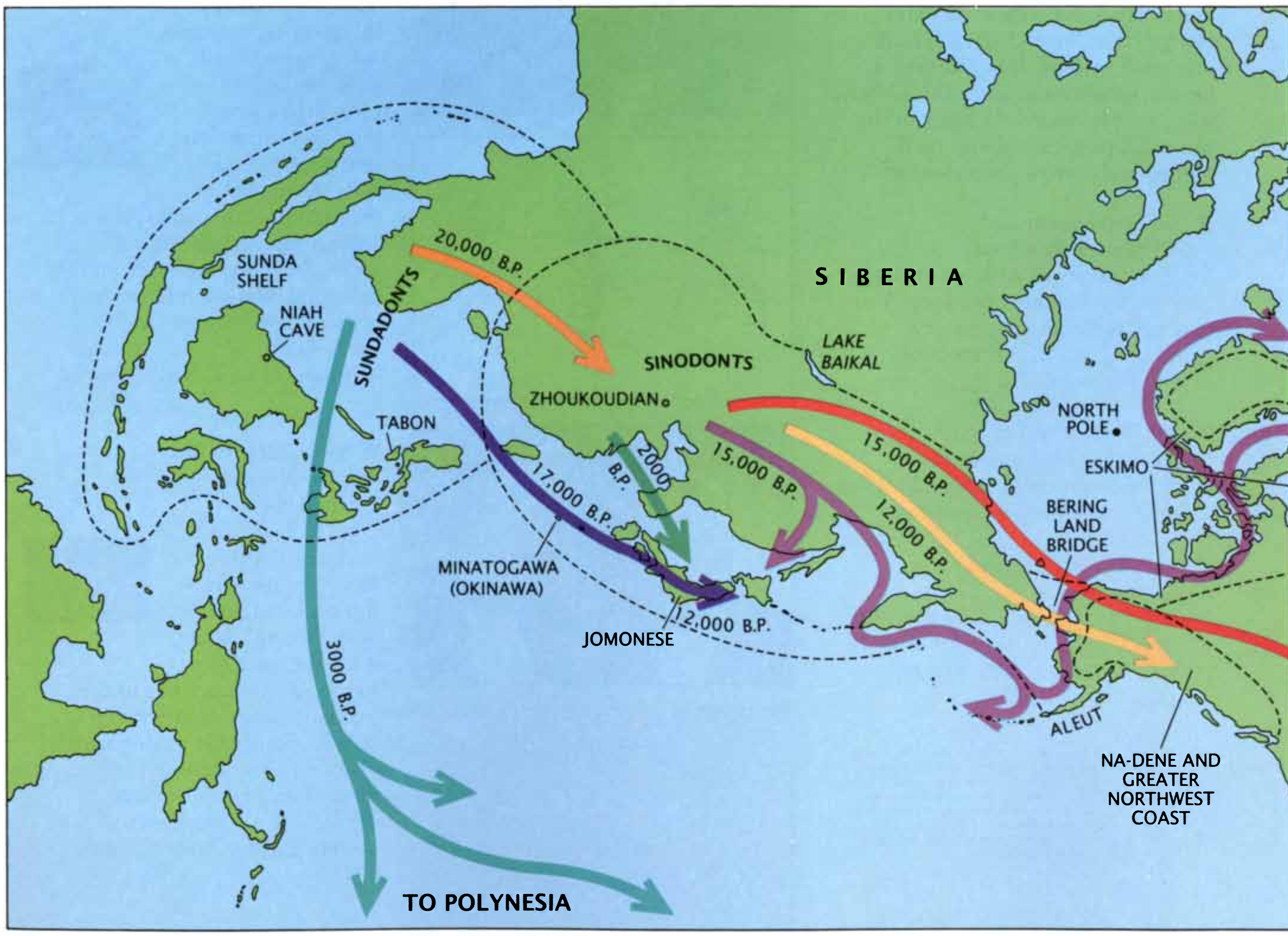
know these voyagers today as Polynesians; they reached points as far east as Easter Island, as far south as New Zealand and as far north as the Hawaiian group. All the human remains recovered so far in these islands have teeth that match the Sundadont pattern of early and present-day southeast Asians.

That the Polynesians had a southeast Asian origin is confirmed by other types of evidence. Extensive linguistic research shows that all Polynesian languages resemble the languages of southeast Asia more than they do any other major language grouping (such as American Indian, northeast Asian, Australian or Melanesian). In addition, the aboriginal peoples of southeast Asia share significant cultural elements with the Polynesians, including pottery, boat shapes and sail forms along with domesticated plants and animals. At the Micronesian Archaeological Conference held in Guam in 1987, studies of cranial shape and tooth size by Michael Pietrusewsky of the University of Hawaii at Manoa, William W. Howells of Harvard University and C. Loring Brace of the University of Michigan buttressed the conclusion still further.

And so it seems that after its inception more than 20,000 years ago, the Sundadont dental pattern expanded rapidly into Japan, then into northeast Asia, where Sinodonty evolved, and much later into the islands of the deep Pacific. What about the later history of the other main grouping, the Sinodonts? They too were an expansive population. The dental story for the peopling of the Americas shows quite clearly that the ancestors of all living Native Americans came from northeast Asia.

Of course, the general conclusion that the first inhabitants of the New World came from northeast Asia is not new, but dental anthropology has been able to supply some novel specificity within that general framework. Convincing archaeological evidence indicates that Siberian Sinodont families moved across the Bering land bridge into Alaska sometime before 12,000 years ago. Some 50 generations later—by 11,000 years ago—they had arrived at the southern tip of Chile, a rate of no more than 10 miles per year on the average.

That the migrating Sinodont population had reached southern Chile by 11,000 B.P. (before the present) has been clearly established by the work of the late Junius B. Bird of the American Museum of Natural History. Bird found teeth and bones of Paleo-Indians who had been intentionally cremated, in association with bones of extinct animals, unquestionable stone



SUNDADONTY AND SINODONTY evolved in east Asia and later spread to the Americas and Polynesia. Sundadonty developed more than 20,000 years ago, when sea levels were lower and the Sunda Shelf occupied the area where Indonesia is today. Some 20,000 years before the present (B.P.) Sundadonts moved northward; Sinodonty developed in northern China and Siberia.

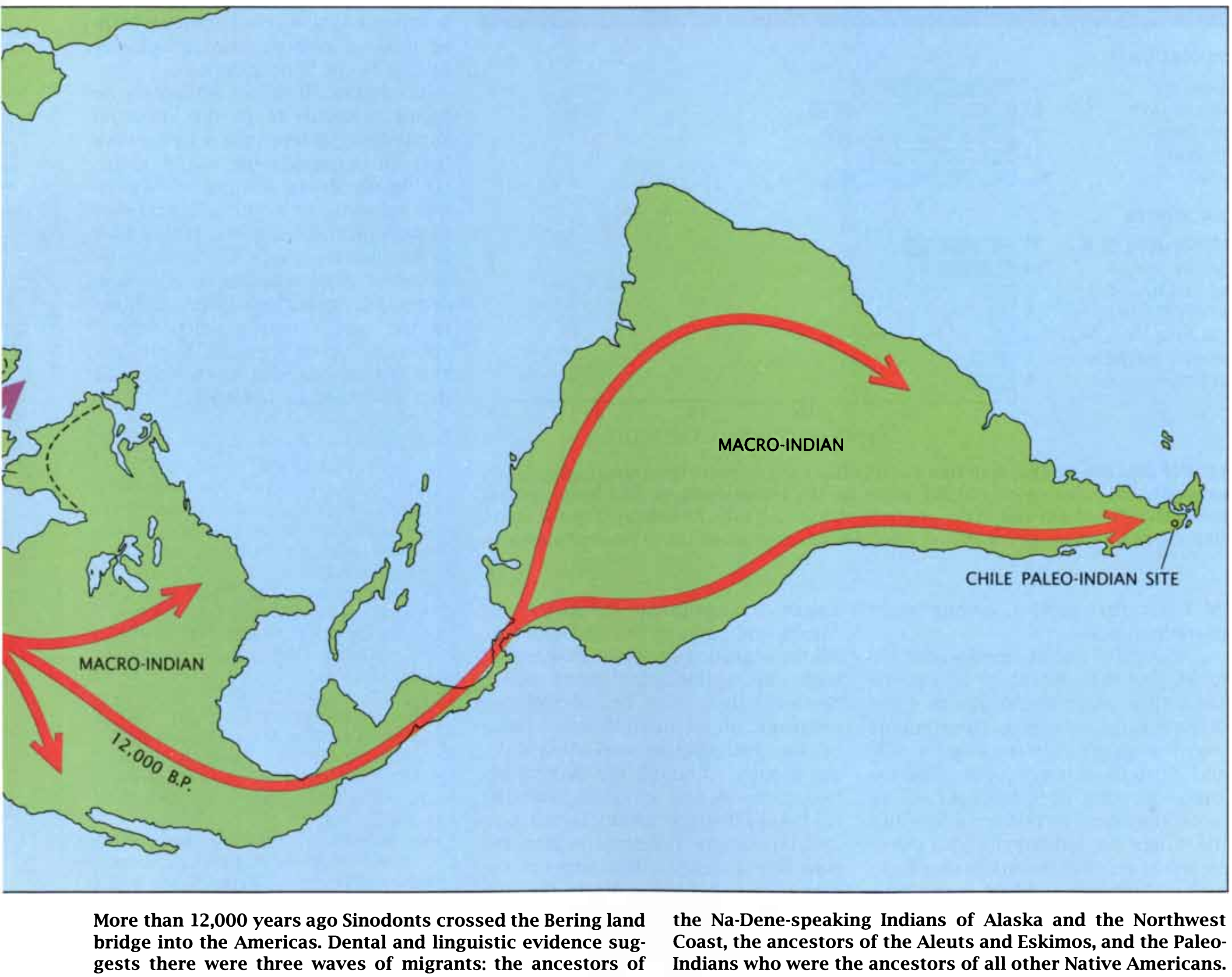
tools and dry grass that had been preserved in caves. This material provided excellent carbon-14 dates. The teeth of these Chilean Paleo-Indians show the unmistakable Sinodont pattern, including one of the hallmarks of Sinodonty: the three-rooted lower first molar.

At least two questions about the great Sinodont migration that have long gone unanswered are now being answered by dental anthropology. The first is: Where precisely within northeast Asia did the migrants come from? Since they were unquestionably Sinodonts, establishing the boundary of Sinodonty would seem to define the region from which the migrating bands were drawn. My recent work suggests that the western border of Sinodonty runs through Lake Baikal and that its southern border is in northern China. This, then, is the area from which the migrants came.

A second question is: How many waves of migrants must be invoked to explain the considerable variation among groups of Native Americans? A spectrum of hypotheses has been proposed, one end formed by the independent proposals of William S. Laughlin of the University of Connecticut (on ecological grounds) and Frederick H. West of the Peabody Museum of Salem in Massachusetts (on an archaeological basis) that there was but a single wave. At the other end of the spectrum, the late Charles F. Voegelin of Indiana University hypothesized on the basis of linguistic considerations that there were hundreds of separate migrating groups.

Two colleagues and I find that an intermediate number of migrations—three—best fits our different types of data. Joseph H. Greenberg is a linguist at Stanford University; Stephen L. Zegura is a physical anthropologist at the University of Arizona. We find that a three-migration model best fits the observed variation in language, genetics and dental patterns and also accommodates a number of major archaeological considerations.

The three waves were probably close together in time, but they yielded groupings that differ greatly in size and geographic extent. One group seems to have included the Paleo-Indians who were ancestral to all South American and most North American Indians. We propose that the single language of these Indians, who were hunters of game on land, evolved into most of the North American and South American Indian languages of today. Teeth indicate the same thing. All these groups have a high frequency of incisor shoveling but a relatively low incidence of three-root-

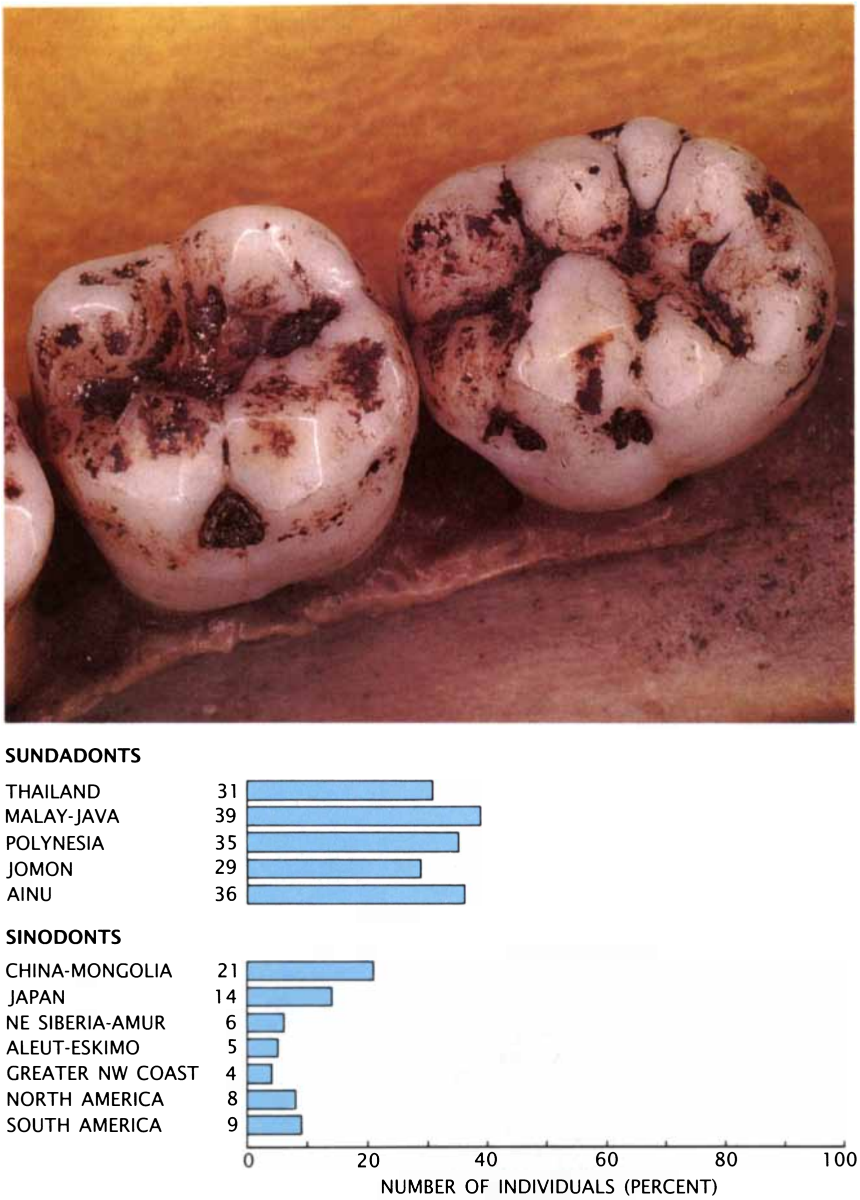


More than 12,000 years ago Sinodonts crossed the Bering land bridge into the Americas. Dental and linguistic evidence suggests there were three waves of migrants: the ancestors of the Na-Dene-speaking Indians of Alaska and the Northwest Coast, the ancestors of the Aleuts and Eskimos, and the Paleo-Indians who were the ancestors of all other Native Americans.

tions among the dizzying variety of peoples, languages and cultures of the Pacific Basin and Rim. A dentally derived concept has been shown to link all this variation into a two-branched system. At its root are the Sundadonts of southeast Asia. After spreading into China and Mongolia 20,000 years ago, Sinodonty developed among the Sundadont population of the northern frontier. From there the new pattern was carried into the Siberian arctic and eventually across the Bering land bridge to North and South America. Much later the Sundadont population expanded again—across Polynesia.

I believe this dental reconstruction explains the main outlines of the peopling of the Pacific Basin and Rim. Yet the reconstruction is by no means complete. There is need for much more information on the teeth of the Australian aboriginals, the Melanesians, the Micronesians and the very early Indonesians. Although these groups must have had ancestors in southeast Asia, it is not known whether those ancestors were all closely related to the Sundadonts.

In addition, it would ultimately be highly desirable to fit the ancestral Sundadont pattern into a framework that encompasses the entire globe. Systematic dental analysis of population movements is still in its infancy as a scientific discipline. It is clear that in the future much work must be aimed at understanding how Sundadonty links up with all the populations to the west: south Asians, central Asians, southwest Asians, Europeans and Africans. At present the pages of that story remain unwritten.



CUSP NUMBER is another trait that distinguishes Sundadonty from Sinodonty. Molars generally have five cusps (raised points on the biting surface). The lower second molar shown in the center of the photograph has only four. As indicated by the chart, four-cusped molars are commoner among Sundadont groups than among Sinodonts.

ed lower first molars, among other characteristics.

A second wave of immigrants entered the New World at about the same time along the southern coast of the Bering land bridge. These people were the ancestors of the linguistically and dentally distinct Aleuts and Eskimos of today. Their frequency of incisor shoveling is relatively low but the number of individuals with three-rooted lower first molars is very high.

The third wave, which may have crossed slightly later than the first two, included the ancestors of the Navaho and Apache along with the Indians who populate the interior of Alaska and parts of British Columbia. All these groups speak languages that make up a linguistic family called Na-Dene; they show dental-trait frequencies intermediate between those of the Paleo-Indian and Aleut-Eskimo groups. Although the overlap between the dental grouping and the Na-Dene language group is not perfect, it is substantial, and considerable New World genetic data support the hypothesis that these are in fact the three units of original migrants.

At the beginning of this article a question was raised about the rela-

FURTHER READING

OUT OF ASIA: PEOPLING THE AMERICAS AND THE PACIFIC. Edited by Robert Kirk and Emöle Szathmary. Journal of Pacific History, Inc., Australian National University, 1985.

DENTOCHRONOLOGICAL SEPARATION ESTIMATES FOR PACIFIC RIM POPULATIONS. Christy G. Turner II in *Science*, Vol. 232, No. 4754, pages 1140–1142; May 30, 1986.

THE SETTLEMENT OF THE AMERICAS: A COMPARISON OF THE LINGUISTIC, DENTAL, AND GENETIC EVIDENCE. Joseph H. Greenberg, Christy G. Turner II and Stephen L. Zegura in *Current Anthropology*, Vol. 27, No. 5, pages 477–497; December, 1986.

LATE PLEISTOCENE AND HOLOCENE POPULATION HISTORY OF EAST ASIA BASED ON DENTAL VARIATION. Christy G. Turner II in *American Journal of Physical Anthropology*, Vol. 73, No. 3, pages 305–321; July, 1987.


Let's get it together... buckle up.
GM
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.
Chevrolet, the Chevrolet emblem and Caprice are registered trademarks of GM Corp. ©1988 GM Corp. All Rights Reserved.

The History of Census Tabulation

By 1880 a growing population and burgeoning data almost crippled the U.S. Census Office. Herman Hollerith came to the rescue in 1890, mechanizing the census with his punched-card tabulating system

by Keith S. Reid-Green

A census taker, parodied in a cartoon that appeared in the *Saturday Evening Post* in 1860, says "I jist want to know how many of yez is deaf, dumb, blind, insane and idiotic—likewise how many convicts there is in the family—what all your ages are, especially the old woman and the young ladies—and how many dollars the old gentleman is worth!"

The questions asked in the early days of the census may have been simpleminded, but when they were asked of every American, the result was an enormous amount of information needing compilation and analysis. The task of processing the data expanded decade by decade as the population grew and the Government sought more information about its citizens. Eventually the information collected by the census created an unparalleled statistical challenge and spawned a revolution in data processing and statistical analysis.

That challenge can best be understood by first describing the methods of data collection and the types of questions that were asked of the populace in the first 100 years of the census. The earliest census takers, called assistant U. S. marshals, sallied forth in August of 1790 to collect information on each American household in accordance with the constitutional mandate in Article I, Section 2: "Representatives and direct Taxes shall be apportioned among the several States . . . according to their respective Numbers. . . . The actual enumeration shall be made within three Years after the first meeting of the Congress of the United States, and within every subsequent Term of Ten Years, in such Manner as they shall by Law direct." As the assistant marshals went from house to house they asked the same questions: the name of the family head and the number of individuals living there who were free white males 16 years and older, free white males under the age of 16, free white females, slaves and other persons.

KEITH S. REID-GREEN is executive director of advanced technology for the Educational Testing Service in Princeton, N.J. He lived in England until the age of 15 and then moved to the U.S., where he received a B.A. at the University of Colorado and an M.S. in computer science at Rutgers University. In 1984 Digital Press published his book *Games Computers Play.* Reid-Green wishes to thank Frederick G. Bohme of the U.S. Bureau of the Census for his help in the preparation of this article.

The answers were then recorded on sheets of paper called schedules, on which a horizontal line represented a single household and vertical columns were for the numbers of persons in each category. When the schedule was completed, the assistant marshals totaled each column and sent the form to the U.S. marshal, who in turn consolidated the data for his district and forwarded that information to the Census Office in Washington, D.C. There workers tabulated and compiled the statistics that poured in from around the country. With each new decade these data processors were faced with an ever larger population and the task of analyzing increasingly complex statistics. In 1790 the population of the U. S. was 3,929,214; by 1840 it had grown to 17 million.

As the demand for information increased, the design of the census form changed to accommodate it. By 1850 each horizontal line on the form represented an individual rather than a household and the entries in each vertical column represented a different statistic for that individual. By 1870 the country had a population of almost 40 million and enumerators were expected to get the following information for every person: name, age, sex, race, occupation, value of real estate, value of personal estate, birthplace, whether parents were foreign-born, month of birth if born within the year, month of marriage if married within the year, school attendance, literacy and whether the individual was deaf and dumb, blind, insane or idiotic. The enumerators were also charged with counting voters (male citizens 21 and over) and people denied the right to vote for reasons other than rebellion. Supplemental schedules listed people who had died during the year or who were paupers or prisoners.

It is hardly surprising that as the census became more complex the incidence of errors—in both data collection and data tabulation—increased. In fact, the situation became so grim that the American Statistical Association petitioned Congress for a revised and corrected edition of the 1840 census on the grounds that there were too many contradictions and unlikely conclusions for the results to be considered valid.

Meanwhile the country's population continued to grow and diversify. By 1880 such great quantities of data were being processed by the Census Office that some of the detailed results of the 1880 census were unavailable until 1888—a full eight years after the data had been collected. It was clear that something had to be done; if nothing was done, the census might have to be abandoned altogether.

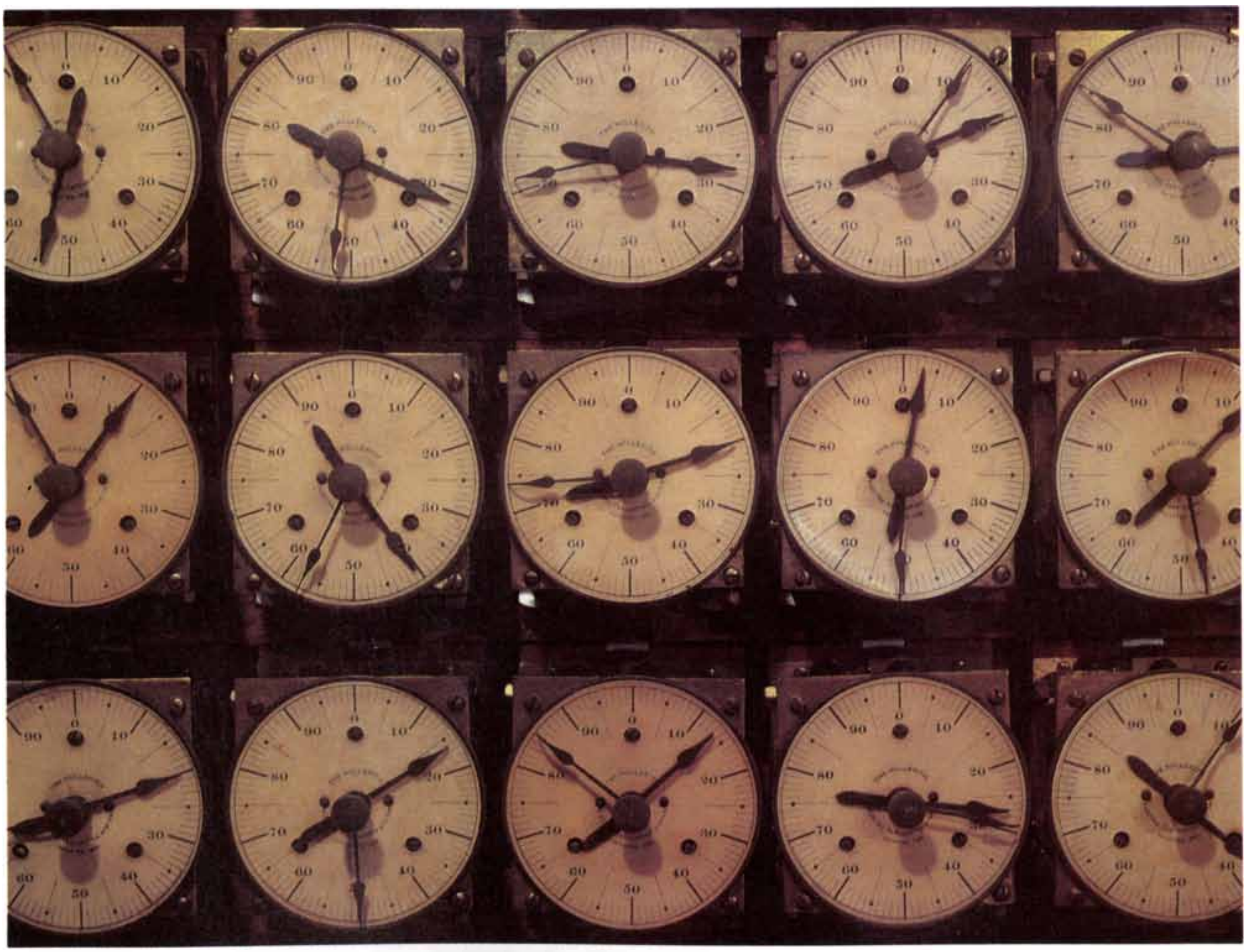
An early attempt to increase the efficiency of data tabulation was made in 1872 by Charles W. Seaton, chief clerk of the census, who invented a simple piece of equipment known as the Seaton Device. It consisted of a wood box into which were set two series of rollers: eight in a row at the top and seven in a row at the bottom. By threading a continuous roll of paper through the box, an operator could position eight lines adjacent to one another (those that needed to be filled in with similar information, such as the age of the head of the household, for example) and thus could record data without having to spend time searching on the spreadsheet for the proper tally spot. The Seaton Device, however, accounted for only one stage in the tabulation process; once the data were recorded, they still had to be summarized by hand and transferred to spreadsheets.

Fortunately 1879 marked a turning point for the Census Office. It was in that year that Herman Hollerith (who later became one of the founders of the International Business Machines Corporation) went to work for the Census Office. Hired to collect statistics on the consumption of steam and water power by the iron and steel industries, Hollerith, who was only 20 years old and a recent graduate of the School of Mines at Columbia University, quickly became familiar with the problems of data processing.

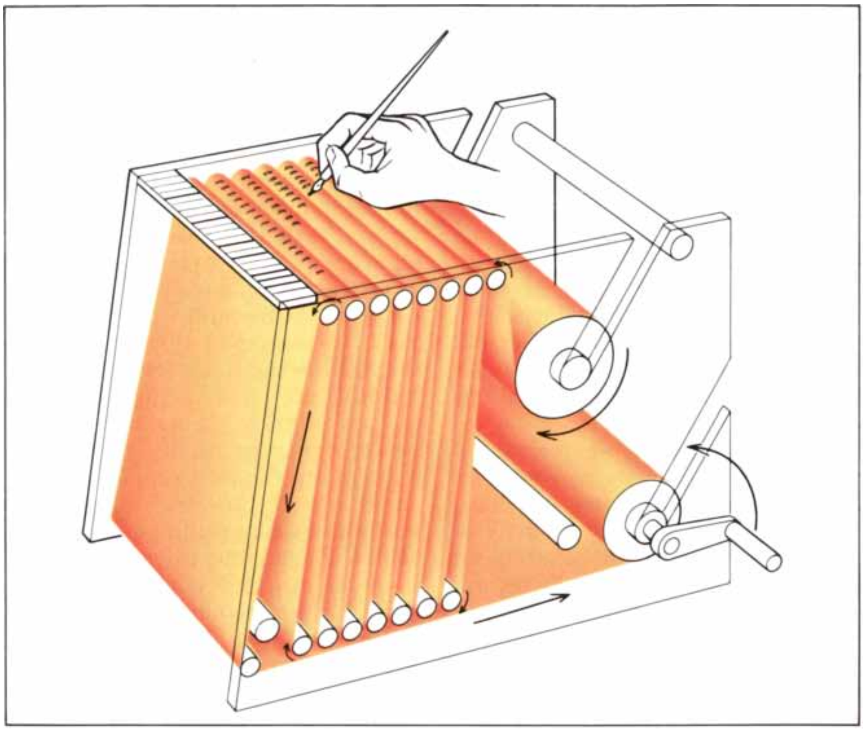
Hollerith's first effort for the Census Office was an electrical tabulating machine that was patented but never actually put into service. Like the Seaton Device, it relied on a roll of paper, but data were represented by the presence or absence of holes. A hole would be punched in a particular spot, for example, depending on whether a person was male or female, native or foreign-born, etc. After one individual had been represented by a cluster of holes, the paper would be advanced and another person would be recorded, and so on until the entire roll had been processed. The paper would then be passed between a metal drum and a set of wire brushes. The brushes would make contact with the drum at each hole, thereby actuating an electromagnet, which in turn would advance a counter. Although such a machine would clearly have been more efficient than anything that existed at the time, Hollerith recognized a major drawback in his design: once the data were tabulated, they would have been almost impossible to retrieve. In order to review the numbers of persons in a category (native-born women in New York, for example) reams of paper would have had to be unwound.

If Hollerith's first attempt at designing an electrical tabulating machine was less than successful, that hardly mattered in the long run, for he was about to perfect the device that would make him famous: the punched-card electrical tabulating machine.

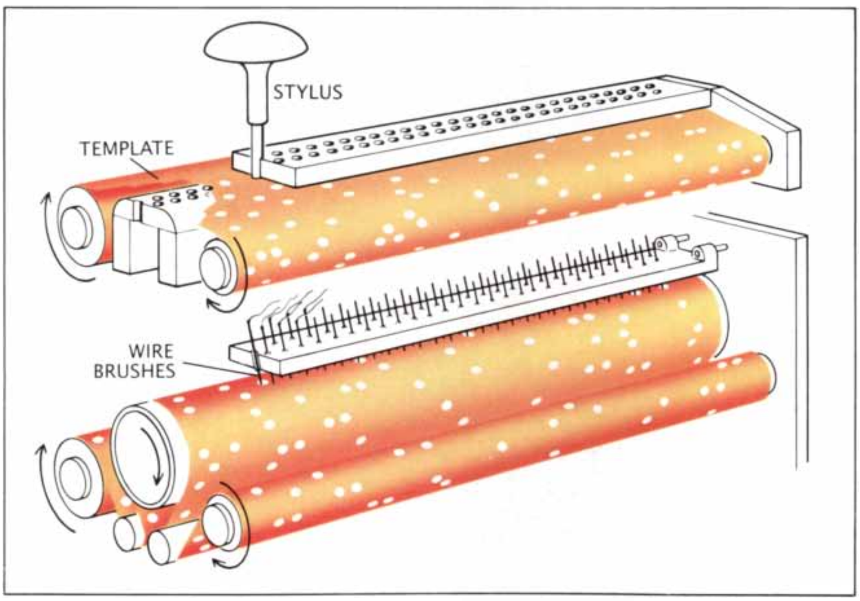
The historical development of the machine can be traced to Hollerith's relation with John Shaw Billings, a physician in charge of compiling health statistics for the Census Office. Ac-



DIAL COUNTERS were part of Hollerith's punched-card tabulating system. Each dial (there were 40) was associated with a specific hole in a punched card; when electrical contact was made through a hole, the dial advanced by one unit. (The large hand represented 10-digit numbers; the small hand, units.) At the day's end the operators recorded the total for each dial.



SEATON DEVICE was an early and largely unsuccessful attempt to speed data tabulation. By threading a long strip of paper through rollers, data from the schedule could be transferred to the paper, where entries in a single category could be aligned next to one another for easier tallying. If, for example, an operator wanted to record the number of children who attended school during the census year, he or she could thread paper through the Seaton Device so that the lines representing education level were brought together at the top of the box. Thus the operator could avoid having to search on a large sheet of paper for the correct entry lines.



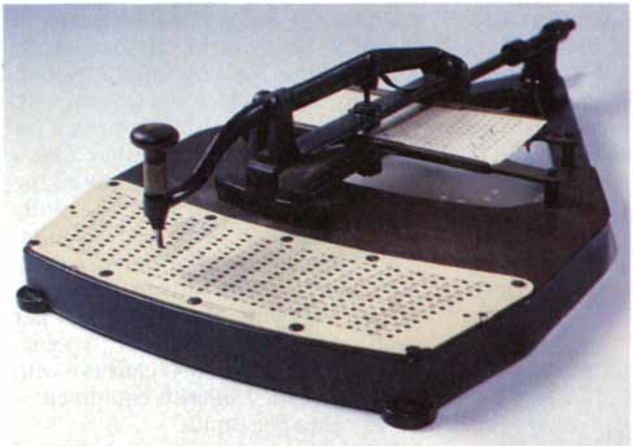
HOLLERITH'S first invention for the Census Office was a two-part device that was patented but never actually put into operation. According to the patent application, a roll of paper would pass under a metal template (*top*) in which there were holes, each one representing a single statistic. The operator would punch holes for each individual according to information on the schedule, then advance the paper, punch another set of holes for another person and so on. The punched paper would then be passed through a counting machine (*bottom*), where it would pass between a metal drum and a series of wire brushes. At each hole a brush would make contact with the drum and advance a counter associated with it by one unit.

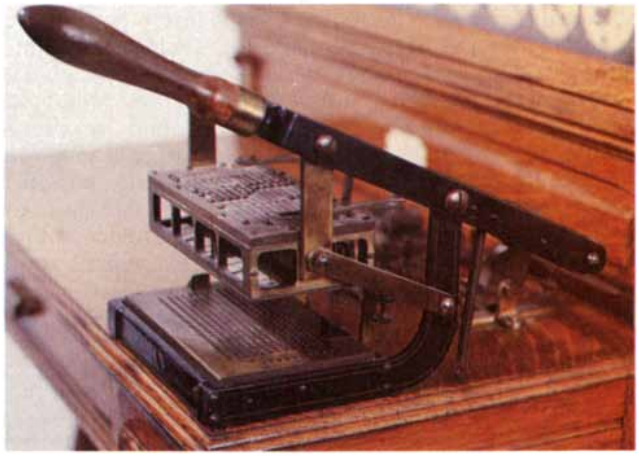
cording to various accounts, while working together on the 1880 census both Hollerith and Billings decided that the 1890 census simply could not be tabulated with existing manual methods. In a 1914 publication Walter F. Willcox (an employee of the Census Office) described overhearing a crucial conversation between the two men: "While the returns of the Tenth (1880) Census were being tabulated at Washington, Billings was walking with a companion [Hollerith] through the office in which hundreds of clerks were engaged in laboriously transferring items of information from the schedules to the record sheets by the slow and heartbreaking method of hand tallying. As they were watching the clerks he said to his companion, 'There ought to be some mechanical way of doing this job, something on the principle of the Jacquard loom, whereby holes in a card regulate the pattern to be woven.'"

Hollerith himself attested to the influence of Billings in a letter dated August 7, 1919: "One Sunday evening, [Dr. Billings] said to me there ought to be a machine for ... tabulating population and similar statistics. We talked the matter over and I remember ... he thought of using cards ... with the description ... shown by notches punched in the edge of the card."

Hollerith, however, is generally credited with being the inventor of the punched-card tabulator. Certainly the concept underlying the system was similar to his first tabulator. Like its predecessor, the new Hollerith machine was a digital device in the sense that information was represented by the presence or absence of holes, but now cards replaced the roll of paper. Each card represented a person and each hole on the card was accorded a specific meaning, corresponding to occupation, education, health or some other characteristic. If the person was a literate citizen in 1890, for example, a hole was punched in column seven, ninth row from the top.

Whatever the origin of the punched-card idea, few would dispute its validity or the flexibility it provided in statistical analysis. The cards made it possible to process information in one way, rearrange it in new combinations and process it again. Cards were developed with holes around their edges; by notching holes with a special punch an operator could effectively remove those holes from the card so that when a knitting needle was inserted through a certain hole—say the one representing the age group from 15 to 20 years old—that subset could be





PANTOGRAPH PUNCH sped the transfer of data from the schedule to a punched card (*left*). When a stylus was inserted into a hole on the template (at the front end of the device), sufficient leverage was created so that a corresponding hole was punched in the card at the other end. Each card represented a person and each hole a different statistic, such as age category or marital status. The press box was an electrical device that counted the holes on a card (*right*). An operator set a card on the perforated base and lowered the upper plate onto the card by depressing the handle. Pins penetrated the card only where holes had been punched; each pin that passed through a hole made electrical contact with a small cup of mercury. When the electrical circuit was completed, the dial counter associated with it advanced by one.

isolated from an entire stack of cards. The operator could then insert the needle through a different hole to determine, for example, how many of those individuals had been educated through sixth grade.

The system by which the data were tabulated was similarly inspired. Once the holes were punched in a card, it was placed in an electrical counter called a press box or pin box. The bottom of the press consisted of a hard rubber plate containing small cups filled with mercury, one for each possible hole in the card, and the top of the press consisted of projecting, spring-actuated pins. As the top of the press was lowered onto a card, the pins that made contact with unperforated card stock were pressed back into their springs. At each hole, however, a pin would pass through the card into a mercury cup, completing an electrical circuit and causing a counter, which had two hands—one for counting units and the other for counting hundreds—to advance by one division. When the card was properly tabulated, a bell rang; failure to ring was an indication that something was wrong with the card. At the end of the day the numbers on the dials were recorded in books and the dials were reset.

Attached to the tabulator was a sorter: a large wood box divided into 26 compartments, each covered by a spring-actuated lid. By means of electrical relays connecting the sorter to the tabulator, the sorter's compartments could be programmed to open to receive cards carrying particular kinds of data. Imagine, for example, that an operator puts the card of a 45-year-old white farm wife into the press box. As the operator closes the press, certain pins make contact with the mercury cups and the counters for white females, females in the 45-to-50 age bracket and occupants of farms of less than 75 acres all advance. If the electrical relays are wired so that the sorter box will respond to age categories, the lid to compartment 10 (for all females 45 to 50) will open. After dropping the card into the open compartment and closing the lid, the operator picks up the next card and slips it into the press box. When all the cards have been sorted, the contents of each box can be counted to determine—in this case—how many females there are in the 45-to-50 category.

Hollerith tested his punched-card system in advance of the 1890 census at Baltimore's Department of Health, where it was set to work tabulating vital statistics for the city. It was lucky that he did so, for he quickly discovered a major flaw in his design. After punching more than 12,000 holes in heavy card stock with a conductor's punch, he lost almost all strength in his hand and wrist.

To overcome the problem he invented the keyboard or pantograph punch, so called because movement at one end of the device was reproduced at the other end. The new punch consisted of a flat frame with a perforated plate at one end and a holder for the card at the other. Above the frame was a swing arm that had a stylus at one end and a punch at the other. When the operator pushed the stylus into a hole on the plate, which contained holes in the same positions as the punched card, the action was transmitted to the other end of the pantograph and a corresponding hole was punched in the card.

The pantograph was an ingenious invention. It not only reduced strain on the operator and increased efficiency but also made it possible to punch holes in a 12-row, 24-column card, thereby increasing the amount of data that could be entered on each card. With the help of the pantograph punch, one operator could punch an estimated 500 cards in one day.

The 1890 census was also expedited by the invention of the four-column gang punch. By pulling a lever on the device, an operator could simultaneously punch one hole in each of the first four columns of a card, which represented geographical data. Because large numbers of cards from the same geographical area were processed together, an operator could set the four-column punched and prepunched cards for each district—as many as five or six cards at a time.

By the end of the 1890 census the Hollerith Electric Tabulating Machine, as it was called, had been significantly refined. As a result efficiency at the Census Office skyrocketed. The average clerk was able to feed almost 8,000 cards per day through the tabulator and many workers did much better. One clerk set a record of 19,071 cards in one day.

Toward the end of the 1900 census an automatic card feeder was added to the tabulator, which eliminated the need to insert one card at a time into the press box. Now operators could process the cards in batches. The efficiency of the automatic tabulator was indeed impressive. A record 84,000 cards were processed in one day, and the average rate was at least six times faster than what the hand tabulator could achieve. In 1901 an automatic sorter was added to the assembly in another step toward increased automation. By pressing a lever an operator could send cards directly from the tabulator into the sorting box.

In 1902 the Census Office was made a permanent branch of the Federal Government and became known officially as the Bureau of the Census in the U.S. Department of Commerce. At that time the cost of the Hollerith equipment (which the Government rented from Hollerith's company for the tabulation period) came under fire. LeGrand Powers, in charge of the 1900 Census of Agriculture, loudly complained that the cost of the automated census was twice that of a manual system. He was, of course, correct: the 1870 census cost 8.77 cents per capita, the 1880 census 11.55 cents per capita and the 1890 census—with the help of the Hollerith equipment—18.44 cents per capita.

Nevertheless, it was widely agreed that going back to manual methods was unthinkable. Instead the Census Bureau chose to reduce costs by constructing its own tabulating equipment. It even made a few improvements in the process. The hand-operated lever on the Hollerith automatic sorter, for example, was replaced with a foot-operated bar that freed the operator's hands, and in 1906 the dial counters were replaced with printing counters that tallied their counts directly on paper.

IMPORTANCE of the Hollerith tabulating equipment to the U.S. Census Office is indicated by its portrayal on the cover of SCIENTIFIC AMERICAN in 1890, the year the equipment was introduced. The various stages in data processing are shown, but not in sequence (*clockwise from upper left*). In the first drawing an operator is seated at the tabulator-sorter, which consisted of a punched-card reader and a sorting box. In the next drawing an operator transfers data from a handwritten schedule to punched cards by means of a pantograph punch. In the next a census worker reads population data from a schedule, adding the numbers on a keyboard punch to obtain a population count, and in the last three drawings bundles of schedules are shown as they arrive at the Census Office from various states and enumeration districts.

By 1930 refinements to data processing were still being developed at the Census Bureau. The pantograph punch was replaced with a key punch (which had previously been adopted to compile agricultural statistics) for the purpose of obtaining a population count. With a card inserted into a movable carriage, an operator needed only to strike a key to make a hole, much as one might hit the keys of a typewriter. Efficiency soared yet again: in less than seven weeks 12.6 million families were counted with the new equipment.

In 1940 the 45-column IBM card was introduced, together with the reproducing punch, a mechanical device that transferred data on one card to one or more additional cards. The IBM printing tabulator, another newcomer in 1940, both summarized and totaled the data from multiple cards and also from various fields on the same card. But the 1940 census is most noteworthy, perhaps, because it was the last one completely processed without electronic computers.

The first computer available commercially, the UNIVAC I, made its debut in the 1950 census, when it tabulated about 20 percent of the punched-card data; the rest of the data were processed on existing tabulating machinery. In 1960 the census entered the computer age wholeheartedly, adopting FOSDIC (Film Optical Sensing De-

HOLLERITH PUNCHED-CARD TABULATING SYSTEM had three parts: a press box, a series of dial counters and a sorting box. An operator would pick up a punched card with her left hand, insert it into the press box and depress the handle with her right hand. When contact was made, the appropriate counters would advance, a bell would ring and the lid of one of the sorting-box compartments would open. The operator would remove the card from the press box with her right hand, drop it into the correct compartment and close the lid, at the same time inserting a new card into the press box with her left hand. The method was remarkably efficient: a single operator could process an average of 8,000 cards per day in this way.

vice for Input to Computers) for tabulating the results. FOSDIC represented a major advance in data processing because it read microfilmed census schedules, eliminating for the first time the need for manual transcription of the schedule data; with FOSDIC, data could be transferred directly to magnetic tapes for processing by the computer.

Subsequent developments proceeded quickly. In 1970 and 1980 information was recorded on 8½-by-11-inch questionnaires that were FOSDIC-scanned at a rate of approximately 240 per minute and the data transmitted by means of tape and/or telephone datalink into the computer's memory. Upcoming changes are likely to further revolutionize data handling by the Census Bureau. In 1990 virtually all of the census will be collected by mail, limiting the need for the old-fashioned door-to-door census taker, except when respondents fail to complete their questionnaires. By the year 2000 written census forms could be eliminated altogether. Instead it may be possible to collect the information by way of interactive cable television, home computers or perhaps even computer-assisted telephone systems. One can therefore say with some assurance that the processing of census data will continue to challenge the ingenuity of computer designers and systems analysts for many years to come.

FURTHER READING

JOHN SHAW BILLINGS, CREATOR OF THE NATIONAL MEDICAL LIBRARY AND ITS CATALOGUE, FIRST DIRECTOR OF THE NEW YORK PUBLIC LIBRARY. Harry Miller Lydenberg. American Library Association, 1924.

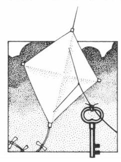
PUNCHED CARD DATA PROCESSING PRINCIPLES. IBM Personal Study Program. International Business Machines Corporation, 1961.

THE DEVELOPMENT OF PUNCH CARD TABULATION IN THE BUREAU OF THE CENSUS, 1890-1940. Leon E. Truesdell. U.S. Government Printing Office, 1965.

HERMAN HOLLERITH: FORGOTTEN GIANT OF INFORMATION PROCESSING. Geoffrey D. Austrian. Columbia University Press, 1982.

THE AMATEUR SCIENTIST

In an emergency stop, should a car's wheels be locked or should the braking be controlled?



by Jearl Walker

A particularly frightening experience when you are driving a car is making an emergency stop to avoid a collision. How should you apply the brakes in order to make the shortest or safest stop? Should you push the brake pedal all the way down and lock the wheels (that is, prevent them from rotating)? Or should you push it down hard but not quite hard enough to lock the wheels? Some driving instructors and a few physicists argue strongly for the first strategy, whereas some physics textbooks argue just as strongly for the second. Indeed, some modern cars have a computer-controlled brake system that automatically adjusts the brakes to keep the wheels from locking. Suppose your car lacks such a system. Which strategy should you follow?

To decide the matter you must consider the frictional force on the wheels from the road. When a wheel rotates smoothly without slipping, the friction is said to be static friction. The size of the frictional force matches the force the wheel exerts parallel to the road. When the car is stationary and exerts no force parallel to the road, there is no friction on a wheel. When the car accelerates and the engine attempts to turn the wheels faster, the driving wheels then push the road toward the rear and the road pushes them forward. It is this forward force on the wheels from the road that propels the car.

When the brakes retard the rotation of the wheels, the wheels push in the forward direction and experience a frictional force toward the rear, which is what slows the car. The upper limit to the force due to static friction is set by the product of the downward force on a wheel and a characteristic of the tire-road interaction called the coefficient of static friction. Normally the downward force in this equation is the amount of the car's weight supported by the wheel. The coefficient of static friction is largely a measure of the small-scale roughness of the tire and the road. A typical value for a dry asphalt road is .8.

If the force on a wheel from the road exceeds the upper limit of the static friction, the wheel begins to slip, and the frictional force on the wheel is then said to be sliding friction. If the brake is applied hard enough, it almost immediately locks the wheel. The frictional force points toward the rear of the car and its magnitude is equal to the product of the downward force on the wheel and a coefficient of sliding friction. Note that although the strength of the force due to static friction can vary from zero up to some maximum amount, sliding friction is set at the size given by the product. (Actually there can be a small variation in sliding friction owing to the speed at which the wheel slips over the road and other factors.)

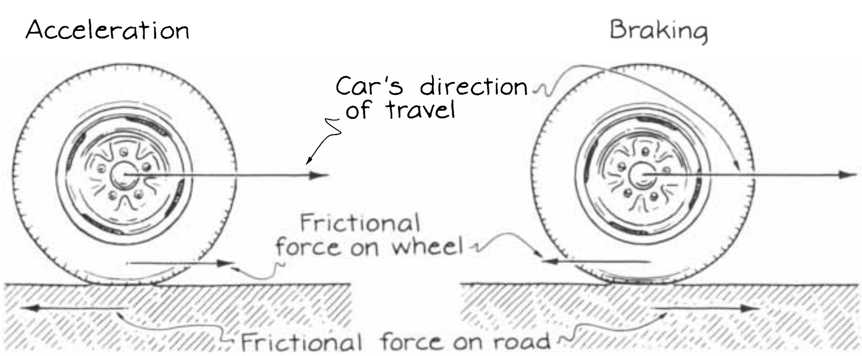
The coefficient of sliding friction is less than the coefficient of static friction. For example, when a standard tire slides over dry asphalt, the coefficient of sliding friction may be .6. The reduction in friction when sliding begins results from several factors, but the most important one involves the heat generated by the sliding. When the road consists of bituminous materials such as asphalt, they melt and a skid mark is left. The tire may also melt. In any case the sliding produces a fluid layer between the wheel and the road that lubricates the motion and reduces the coefficient of friction.

And so should you lock the wheels in an emergency braking or not? If you brake hard without locking the wheels, the static friction on the wheels can be as large as its upper limit. If instead you lock the wheels and they begin to slide, the friction is smaller than that upper limit because of the lubrication. Since in each case it is friction that brings the car to a stop, the best decision seems to be the first one, in which case the friction is larger and the car should stop in the shortest distance. Thus state some physics textbooks.

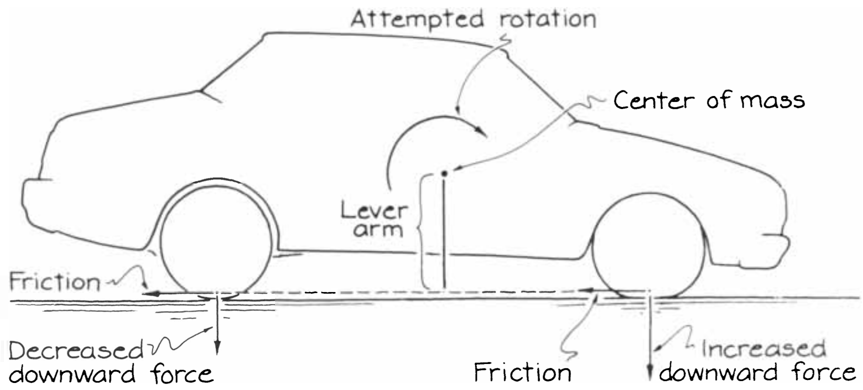
Driving experts sometimes differ with the conclusion. At least one has argued that when you are in an emergency braking situation, you surely do not have time to find the braking that maximizes the static friction while avoiding any locking of the wheels. The point is certainly correct: if you adjust and then readjust the braking, the time required must inevitably add to the stopping distance.

Suppose you ignore that practical point. How then do you answer the question about braking methods? The matter was taken up in 1979 by Daniel P. Whitmire and Timothy J. Alleman, who were then at the University of Southwestern Louisiana. They pointed out that, contrary to the standard textbook argument, experiments reveal that a stop in which all four wheels are locked usually requires less distance than a stop without any sliding. Their key point is that during braking, forces due to friction create torques on the car that significantly change the downward forces on the wheels and hence the frictional force.

To understand their point, consider



Friction on wheel and road



Torque created by friction during braking

a car that is moving forward when the brakes are applied [*see illustration at right*]. If none of the wheels slides, then each experiences a static frictional force toward the rear that slows the car. The forces also create torques that attempt to rotate the car vertically around its center of mass, so that the rear rides up and the front is pushed down. For each wheel the torque is equal to the product of the friction and a certain distance that is called a lever arm. To find the lever arm, mentally extend a frictional force under the car. The lever arm is a line perpendicular to the extension that passes through the car's center of mass. Since the friction is at road level, the length of a lever arm is the height of the center of mass. You can determine the direction of rotation associated with a torque by imagining that the force attempts to rotate the lever arm itself around the center of mass.

During hard braking the rotation of the car may be felt if the car's suspension system yields sufficiently: the car pitches forward, increasing the weight on the front wheels and decreasing the weight on the rear wheels. Even if the car were rigid, and so unable to rotate, the impulse to rotate would produce the same result: the torques would relieve some of the downward force on the rear wheels and increase the downward force on the front wheels. Hence the upper limit of the static friction becomes larger on the front wheels and smaller on the rear wheels than before the braking began.

To see how these changes figure in braking, Whitmire and Alleman considered several situations. For comparison they employed the textbook example in which the torques are not considered and each wheel is assumed to experience the same upper limit to the static friction. Let D be the stopping distance under these circumstances. Next they considered the action of the torques. Suppose that the brakes are identical and the driver has got the rear tires on the verge of slipping. The friction on the rear wheels is then at the upper limit for static friction, but the upper limit has been diminished by the torques. Since the braking on the front and the rear wheels is assumed to be identical, the front wheels must experience the same amount of friction. With the friction small all around, the stopping distance may be as large as 1.5 D.

Such controlled braking may actually be less effective at stopping a car than a full slide. The stopping distance for a fully sliding car is not altered by the torques, because the reduction of the friction on the rear wheels is exactly matched by the increase of the friction on the front wheels, and so the combined friction is the same as if there were no torques. As a result, if the coefficient of sliding friction is 20 percent less than the coefficient of static friction, the stopping distance for a fully sliding car is 1.25 D, which may be appreciably better than the distance for controlled braking.

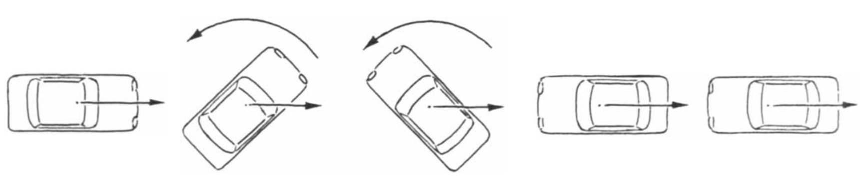
The stopping distance for controlled braking can be reduced if the brakes are applied firmly enough so that the rear wheels begin to slide and the front wheels are on the verge of sliding. Then the combination of the maximum static friction on the front wheels (which is large because of the torques) and the sliding friction on the rear wheels stops the car in a distance that is only slightly longer than D and hence is smaller than the stopping distance for a fully sliding car.

The brakes on cars are often biased to counteract the torques created by the friction: when you press down on the brake pedal, you engage the front brakes more than the rear brakes. Suppose the bias is so strong that the front wheels are put on the verge of sliding while the rear wheels are far from the sliding limit. Then the front wheels experience the maximum static friction while the rear wheels may experience only a small static friction, and again the stopping distance can be dangerously larger than D, and even larger than the distance for a fully sliding car.

The situation improves if the brakes are applied hard enough so that the front wheels begin to slide and the rear wheels are put on the verge of sliding. Then the front wheels experience sliding friction while the rear wheels experience the maximum static friction, and the stopping distance is only slightly larger than D.

There is one more condition to consider. The brake bias can be adjusted so that front and back wheels can be simultaneously put on the verge of sliding and all the wheels experience the maximum static friction. Only in this ideal case is the stopping distance D, as predicted by a textbook calculation. Such adjustment, however, may not be practical because it depends on the coefficient of static friction. Although the bias can be optimized for one type of road surface, it would then not be optimum for a different type of road surface with which the tires experience a different coefficient of static friction. (Here one sees the rationale for computer-controlled brake systems: the bias can be automatically adjusted for any road.)

In summary, then, locking all four wheels and sliding may be the best strategy in some emergency stops if the idea is to minimize the stopping distance. Its advantage is greatest if the torques on a particular car tend to put the rear wheels on the verge of sliding before the front wheels reach that point. The strategy does have one serious flaw, however: if the car is fully sliding, you have virtually no con-



A spin

trol over where it is headed, because of the lubricating fluid between the tire and the road. If the car happens to be turning when it begins to slide, you might quickly spin out of your lane of traffic. Spin can also be initiated if the wheels do not experience identical sliding friction or if the road is tilted or crowned, as it often is to improve drainage. A stop with the car both sliding and spinning might prove to be even more dangerous than a controlled stop, even if the controlled stop requires a greater distance.

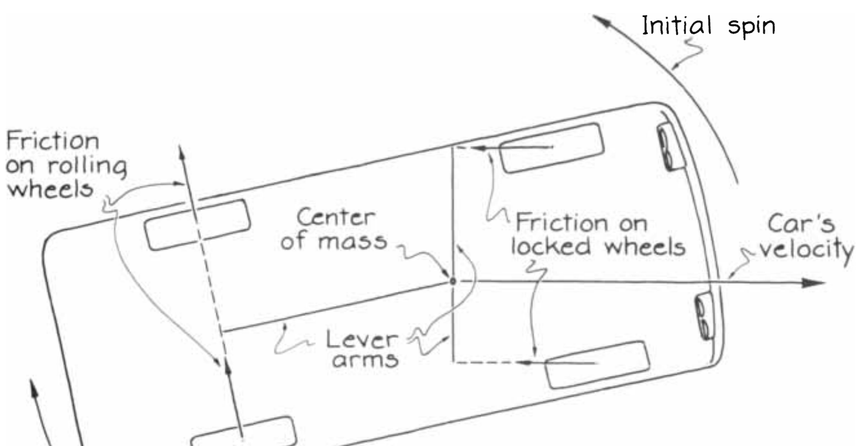
The possibility of "spinout" was investigated in 1984 by William G. Unruh of the University of British Columbia for situations in which only the front wheels or only the rear wheels begin to slide. As you may have noticed in driving emergencies, if only the front wheels lock, the car is often stable in the sense that it continues to point forward, but if only the rear wheels lock, the car is likely to spin around until it points toward the rear and travels backward down the road.

To follow a simple version of Unruh's analysis (he also offered a highly refined version), consider a car that suddenly begins to spin counterclockwise [*see illustration below*]. Suppose that the front wheels are locked and the rear wheels are still rolling. Since the car is spinning, the rear wheels must be sliding sideways and experiencing a sliding frictional force that is parallel to the rear axle. The front wheels also experience sliding friction, but because they are locked, the forces point directly backward, opposite to the intended direction of travel.

Each frictional force generates a torque that tends to rotate the car horizontally about its center of mass. To evaluate the torques, take an overhead view of the car and ignore the height of the car's center of mass above the road. Then, to find the lever arms associated with the torques, extend an arrow representing each force until it intersects a perpendicular line that goes through the car's center of mass. The length of the perpendicular line is the lever arm, and the direction of the rotation caused by a torque can be inferred by imagining that the force tends to rotate the lever arm itself about the center of mass.

When the front of the car has just begun to spin to the left, the approximately equal torques from the front wheels attempt rotations in opposite directions and cancel, but the torques from the rear wheels both promote a clockwise rotation that opposes the spin. Even if the car turns appreciably, so that the torque from the left front wheel increases because of a lengthened lever arm, the torques due to the friction on the rear wheels continue to dominate because of their longer lever arms, and eventually the spin is halted and then reversed until the car is again pointed in the proper direction. Hence when only the front wheels lock, any chance rotation of the car is automatically corrected by the friction on the rear wheels.

Suppose instead that only the rear wheels are locked. They experience frictional forces that are directed backward, while the still rolling front wheels experience frictional forces that are parallel to the front axle [*see illustration on opposite page*]. Early in the spin the torques from the rear wheels essentially cancel, because they attempt rotation in opposite directions. The torques from the front wheels do not cancel, because they both promote a counterclockwise rotation, and so the car continues to spin. As the car's misalignment increases, the torque from the right rear wheel increases because of its lengthened lever arm, but it is still no match for the combined torques from the front wheels. And so the spin goes unchecked and the car turns completely around until it is traveling down the road backward.



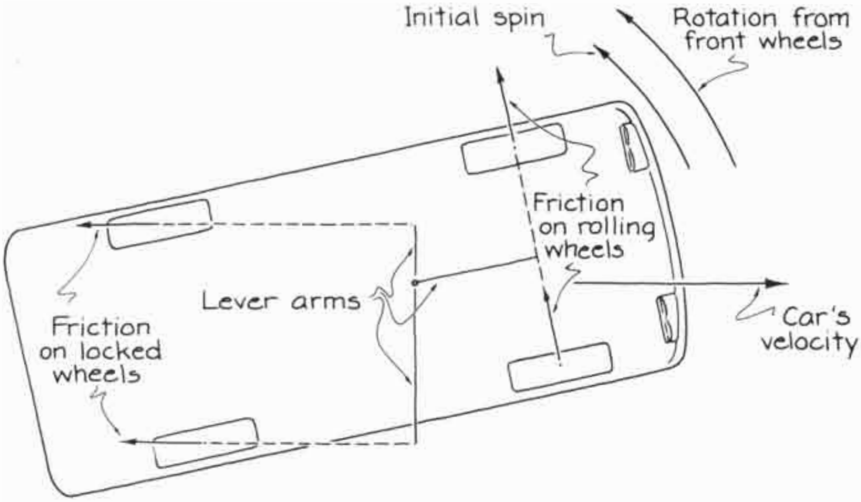
A car with locked front wheels begins to spin

Unruh also considered how the position of a car's center of mass influences the likelihood of such a reversal. If the engine is in the front of the car, the center of mass is forward of the car's midpoint and less weight rides on the rear wheels than on the front ones. The rear wheels are then more likely to lock first during an emergency stop, sending the car into a spin.

When the engine is in the back of the car, as it was in a Volkswagen "Beetle" once owned by Unruh, the car is much stabler against spin. Then the center of mass is behind the car's midpoint and more weight rides on the rear wheels than on the front ones. The front wheels are more likely to lock first, and so any spin is quickly corrected by the friction on the rear wheels. Some drivers arrange for such stability even in a car where the engine is mounted in the front—they pack bags of sand or some other heavy objects in the back of the car. The precaution might indeed help a driver who faces the possibility of braking on an icy street where otherwise the rear wheels would easily become locked.

Two initial conditions affect the possibility that a car with a front-mounted engine will spin when its rear wheels are locked. If the car is suddenly at an angle to the intended direction of travel but has no spin yet, it will begin to spin out of control if the angle is larger than some critical value. The critical value is small for most street speeds, and it gets even smaller for progressively higher speeds. If instead the car is initially pointed in the intended direction and is given some spin by, say, a nonuniform part of the road, it will continue to spin if the initial spin is larger than some small critical value. Again speed works against the driver. In most practical situations, particularly at high speed, the locking of the rear wheels is certain to send the car into an uncontrollable spin.

The existence of a critical angle and a critical initial spin has to do with the fact that the frictional forces not only



A car with locked rear wheels begins to spin

create torques on the car but also slow it and tend to change its direction of travel. If neither the critical angle nor the critical initial spin has been exceeded, the action of the forces actually stabilizes the car. For example, if the car is suddenly turned slightly leftward, the friction that then works on the front wheels might accelerate the car's center of mass to the left, thereby bringing its direction of travel in line with the direction in which it is pointing before the spin can build.

You might be able to correct a spin by turning the front wheels toward the intended direction of travel; at least, that is what is commonly advised. If the speed is low, the scheme should work even if you turn the wheels somewhat too far and then have to turn them back. If the speed is high, however, any error on your first try will shoot the car through the proper orientation and spin it out of control in the opposite direction.

All the braking schemes I have presented ignore a practical aspect in an emergency: reaction time. Since you cannot recognize a dangerous situation and apply the brakes instantaneously, you must necessarily travel down the road a certain distance before the braking begins. Some legal textbooks that deal with traffic accidents place lower limits on reaction times. If the danger is recognized quickly and calls for nothing more complex than braking, you might need only a quarter of a second to perceive the danger and then another quarter of a second to apply the brake. Suppose that you are traveling at 90 kilometers per hour (about 55 miles per hour) when the emergency arises. Then the minimum distance based on your reaction to the danger and full application of the brake is about 13 meters. Of course, if the danger is difficult to ascertain or if you are distracted momentarily or slowed by alcohol, the distance associated with your reaction can be much longer.

There is another factor in an emergency stop—one that you are pretty well stuck with. It is the coefficient of friction, either static or sliding, between the wheels and the roadway. Extensive experiments have shown that the coefficient can vary considerably depending on the type of tire and the degree to which it is inflated. The coefficient can also vary with the type of paving material, how long the pavement has been in place and how heavily traveled the road has been. For example, a concrete road polished by heavy use can have a coefficient of sliding friction that is only 70 percent of what it was when the surface was new, and of course the decreased coefficient considerably lengthens the minimum stopping distance.

FURTHER READING

THE PHYSICS OF TRAFFIC ACCIDENTS. Peter Knight in *Physics Education,* Vol. 10, No. 1, pages 30-35; January, 1975.

EFFECT OF WEIGHT TRANSFER ON A VEHICLE'S STOPPING DISTANCE. Daniel P. Whitmire and Timothy J. Alleman in *American Journal of Physics,* Vol. 47, No. 1, pages 89-92; January, 1979.

AUTOMOBILE STOPPING DISTANCES. L. J. Logue in *The Physics Teacher,* Vol. 17, No. 5, pages 318-320; May, 1979.

INSTABILITY IN AUTOMOBILE BRAKING. W. G. Unruh in *American Journal of Physics,* Vol. 52, No. 10, pages 903-909; October, 1984.

TRAFFIC ACCIDENT INVESTIGATOR'S MANUAL FOR POLICE. J. Stannard Baker and Lynn B. Fricke. The Traffic Institute, Northwestern University, 1986.

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.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.25 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/89. $1.25 + 0.00.

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. The date of the last issue on subscriptions appears at the right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name

New Address

Street

City

State and ZIP

Old Address

Street

City

State and ZIP

COMPUTER RECREATIONS

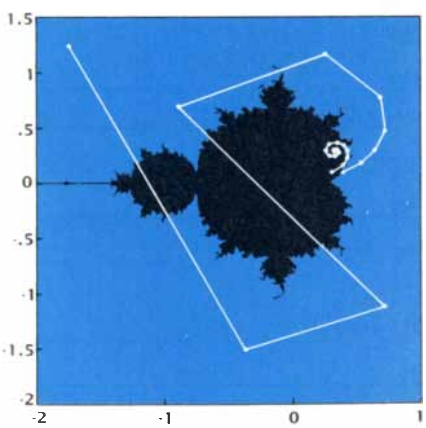
A tour of the Mandelbrot set aboard the Mandelbus

by A. K. Dewdney

Not Art and Science serve, alone;
Patience must in the work be shown.
A quiet spirit plods and plods
  at length;
Nothing but time can give the brew
  its strength.
—JOHANN WOLFGANG VON GOETHE,
Faust

Art and science seem to blend in the Mandelbrot set. Its astonishing complexity symbolizes the growing field of chaos and nonlinear dynamics, and yet even people who have no idea of the set's real physical significance find a strange beauty in its murky interior and bejeweled halo. And, as Goethe's insightful passage reminds us, patience is indeed required when one sets out to "brew" an image of the Mandelbrot set. Although a personal computer can carry out the necessary calculations thousands of times a second, it may take up to several hours for the computer to deliver a finished image.

The Mandelbrot set (named after its discoverer, Benoit B. Mandelbrot of IBM's Thomas J. Watson Research Center in Yorktown Heights, N.Y.) actually consists of an uncountable infinity of points. The images in this month's column are constructed by determining which points in a smallish collection of points (a mere few thousand or so) are members of the set. Each pixel, or picture element, in an image corresponds to a point in the collection. Each image therefore represents one part or another of the set magnified to some degree and colored according to the taste of the Mandelbrot mappers featured here. Looking at the pictures, readers may well wonder: What restless beauty stirs within the Mandelbrot set? What process creates these extraordinary forms, this complexity that plumbs the very depths of infinity? The simplicity of the answer stands in stark contrast to the complexity of the set.



Two possible Mandelbus routes

In two earlier columns I described in some detail how to generate the Mandelbrot set by computer [see SCIENTIFIC AMERICAN, August, 1985, and November, 1987]. This time around I want to ensure that even readers who suffer from acute math anxiety gain some insight into the set. Explaining to certain people that the Mandelbrot set has something to do with "complex numbers" causes them to blanch and mumble excuses about having to be somewhere soon. Fortunately it is possible to avoid the subject of complex numbers altogether. I invite those readers (and all the rest, of course) to accompany me to an imaginary plane where the Mandelbus will take us on a special tour of the Mandelbrot set.

Like points on the earth's surface, which can be specified in terms of a latitude and a longitude, the points on the plane also have coordinates. The point (0,0), called the origin, lies at the center of the plane. With a stroke of a magic wand I convert the reader into an unimaginably tiny being sitting on a point whose coordinates are (a,b). Where exactly is that? Why, anywhere the reader wants to be. But once the decision is made, we shall agree that a and b will have specific values, such as 2.78 and –.43, in which case the reader would be 2.78 units north and .43 unit east of the origin. What is a unit? It makes no difference whether one thinks of it as a mile or a meter. After all, any distance seems enormous when one is as small as a point.

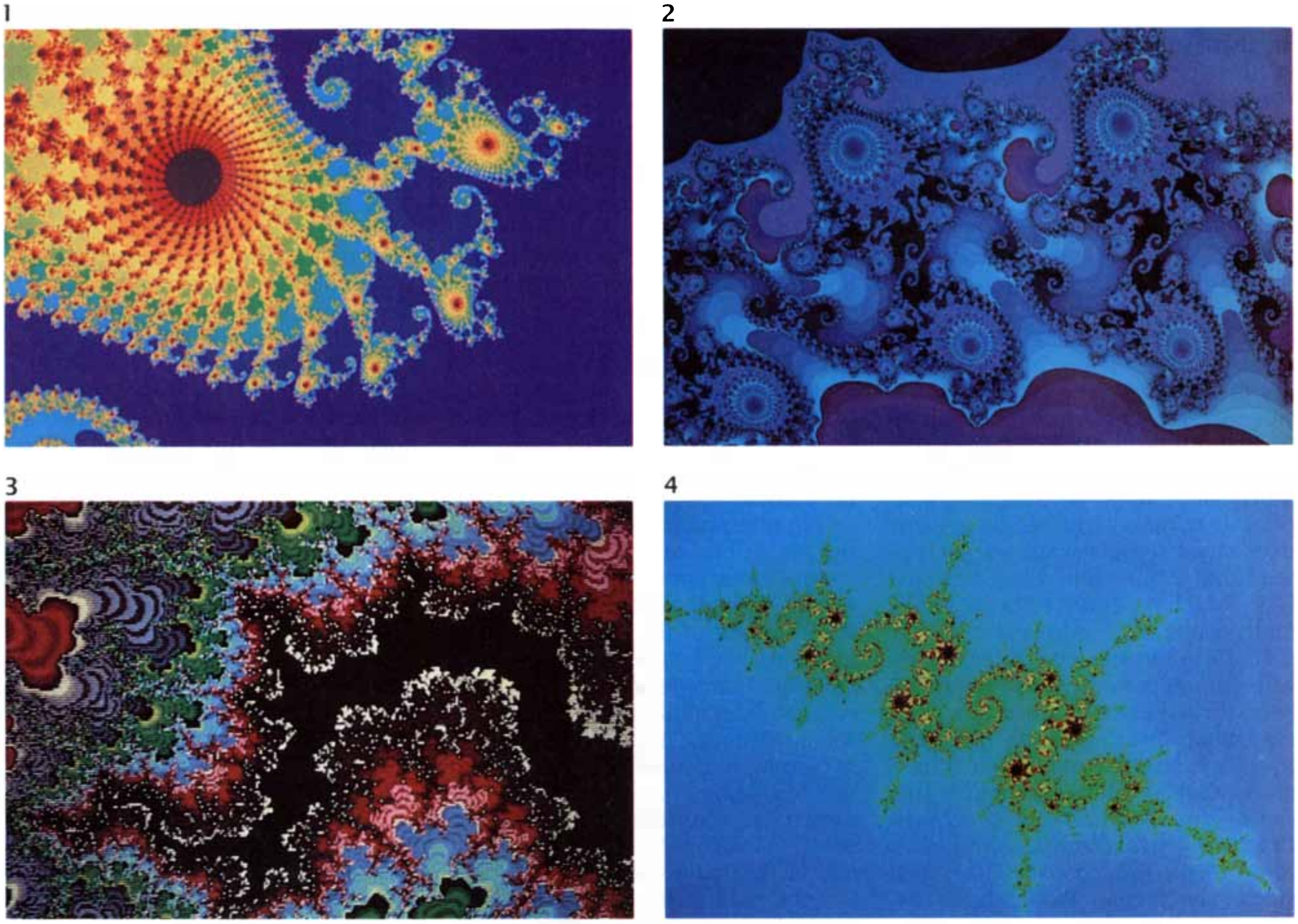
Now that the cooperative reader—tiny and alone—sits on a point on the plane, he or she is ready to catch the Mandelbus. Before the bus arrives I must explain its route. It begins its journey at (0,0) and then goes directly to (a,b) to pick up the reader. The ride will be solitary, for no more readers will be picked up as the Mandelbus careens from point to point, making an infinite succession of stops. To be specific, the Mandelbus route is given by a relatively simple formula that tells the driver what point to visit next after stopping at (x,y). It is

$$(x,y)^2 + (a,b)\,.$$

What does the formula mean? The term $(x,y)^2$ is merely shorthand for the point having coordinates $x^2 - y^2$ and $2xy$. Adding (a,b) to $(x,y)^2$ means one has to add a to the first coordinate and b to the second. At the risk of becoming tedious, the coordinates that result from all this arithmetic are therefore $(x^2 - y^2 + a,\ 2xy + b)$. Where is this? To find out, substitute into the formula the coordinates x and y of the previous stop as well as the coordinates a and b of the first stop.

The Mandelbus always starts at the origin, and from there it zips to (a,b). To see this the reader has only to substitute 0 for both x and y in the formula above. All the squares and products of x and y vanish, leaving only a and b as the new coordinate positions. To see where the Mandelbus goes next, one merely substitutes the coordinates a and b for x and y. When that is done, one sees that the second stop is always $(a^2 - b^2 + a,\ 2ab + b)$.

It is a jerky, bone-jarring ride on the Mandelbus. To give the reader some idea of what it is like, an aerial view of part of two possible rides is shown in the illustration on this page. In one case the bus leaves the origin, stops to pick up the reader and then visits a strange series of points that seem to spiral back to the starting point. In the other case the reader visits a sequence of stops that are at first close together but gradually become spaced farther apart before they head for infinity. It might be a good idea to get off at the



A collection of Mandelbrot images

next stop if you happen to be on that particular trip.

Although the two trips are markedly different in character, each was determined entirely by the first stop (a,b) according to the formula. In the first case, for example, the reader gets on the Mandelbus at (.300, .100). When those coordinates are substituted into the formula, the coordinates (.380, .160) of the second stop are derived. To gain some confidence in Mandelbrot mechanics it would not be a bad idea to take a calculator and actually work out the coordinates of the third stop by substituting .380 for x, .160 for y, .300 for a and .100 for b in the formula. The answer is (.419, .222). This simple, repetitive arithmetic powers the Mandelbus indefinitely.

Although the journey can last for as long as one is willing to repeat the calculation, the route the Mandelbus actually follows will be either bounded or unbounded. By this I mean that the succession of stops either will be confined forever within a zone around the origin (as in the first case) or will eventually escape the zone and head out to infinity (as in the second case).

What, then, is the Mandelbrot set? It is simply the set of all the points (a,b) that result in bounded Mandelbus routes. The set is a solid continuum in the middle of the plane, but it sends out filaments into its surroundings in highly complicated and subtle ways. Computer programs that generate images based on the Mandelbrot set decide the color of a pixel on the screen by determining whether the Mandelbus roars off to infinity from the point (a,b) represented by that pixel and, if so, how quickly.

Some readers may be most comfortable riding the bounded routes that start within the Mandelbrot set; they shudder at the thought of the trip to infinity. The prospect of taking only bounded trips, however, seems to depress mathematicians. That is probably why they color the set black.

It is a proved theorem (and therefore undeniably true) that if the Mandelbus ever reaches a point two or more units from the origin, it is destined for infinity. In such a case its first stop (a,b) lies outside the Mandelbrot set. But how many stops does an infinity-bound Mandelbus need to make before it is at least two units away from the origin? That number is called the dwell of (a,b).

When one attempts to compute the dwell of (a,b), one never quite knows when to stop. The point might result in 100 or 1,000 stops before the Mandelbus leaves the scene for good. On the other hand, the point might be inside the set, in which case no amount of waiting justifies any color but black for its corresponding pixel. Near the set's boundary are points that have very high dwell values. For such points it may take a million iterations before the Mandelbus equals or exceeds the critical distance of two units from the origin. A program must therefore make a more or less arbitrary decision about when to stop iterating the formula for a given point (a,b). Limits as low as 100 or even 50 iterations produce beautiful and reasonably accurate images of the set, in spite of the fact that some pixels will be colored black erroneously.

The assignment of other colors to a pixel is also based on the dwell of the point corresponding to the pixel. The programmer might decide, for example, to color pixels either black, violet or chartreuse depending on whether their corresponding points have respectively a dwell value of 100 or more, between 50 and 99 or between 49 and 1. The colors and the ranges of dwell to which they apply are arbitrary, yet they are choices that can result in breathtaking beauty (as is evident in the collection of images in the illustrations on pages 109-111) or aesthetic disaster (as I suspect my color-assignment example would be). Indeed, half of the programmer's task lies in assigning colors to points outside the set; the other half lies in finding interesting areas of the plane to map. Actually one is affected by the other, since it is the assignment of color that makes visible the regions that have been given names such as Sea-Horse Valley and the Uttermost West (which is populated by topknots, midgets and other odd creatures).

To fill an entire computer screen with colored members and nonmembers of the Mandelbrot set, one must systematically assign a value of *a* and *b* to each pixel on the screen from a given range of values for *a* and *b*. This can be done by means of two loops. One loop varies *a* from one end of its range to the other in, say, 200 equal steps. The second loop, which contains the first, varies *b* in as many steps from one end of its range to the other. The number of steps for each loop reflects the number of pixels one's display screen has in the horizontal dimension (for the values of *a*) and in the vertical dimension (for the values of *b*).

Because each image contains only a finite sampling of points in a particular area of the Mandelbrot set, it can never show all the detail that is actually there. But by making the ranges of *a* and *b* narrower one can increase the level of "magnification." Indeed, the Mandelbrot set has detail at all levels of magnification. For this reason one of the most popular games Mandelbrot enthusiasts play is magnifying areas of the set until they exhaust the capability of their hardware, software or patience.

The magnification of a particular image can be calculated as follows. Suppose one computes an image of a square region of the plane that has a length *s* on a side. The magnification of the displayed image will then be 1/*s*. A square image that is .02 unit wide, for example, will have a magnification of 50 when it is displayed on the screen.

I am often tempted to climb aboard the Mandelbus myself and switch its control from automatic to manual. It would be high adventure indeed to drive up one of the Mandelbrot set's filaments on a sightseeing tour. Far from the Mandelbrot set, for example, one can find miniature copies of the set, apparently floating on the plane. According to mathematicians, the entire Mandelbrot set actually is connected, so that one might think it possible to drive the Mandelbus out to one of the "mini Mandelbrot" copies. But the road is like no road on the earth. If one enlarges repeatedly almost any of the filaments that surround the Mandelbrot set, one finds only a series of tiny black islands that are seemingly disconnected. To be sure, there are smaller islands between larger islands, but a continuous road rarely if ever shows up.

Now I shall turn the tour over to some readers who, after long hours of exploration and aesthetic fine-tuning, have created exemplary renderings of the set. Walter S. Strickler of Boulder, Colo., provides image 1. The image, which he calls Peacock, was created by a program he wrote for his Amiga computer. The *a* values in the image range from –.750 to –.746, and the *b* values range from .0986 to .1014. The magnification of Strickler's image can be calculated according to the simple formula I mentioned above. The smaller of the two sides of the rectangular area covered by the image has a length of .0028. Its inverse (1/.0028) yields an approximate magnification of 357.

Images 2 through 4 are actually slide benefits of my subscription to *Amygdala,* a newsletter published by Rollo Silver in San Cristobal, N.Mex. (One can subscribe to the newsletter or order the full set of images on transparencies by writing to Silver at Box 219, San Cristobal, N.Mex., 87564. Correspondents are asked to send a self-addressed, stamped envelope.)

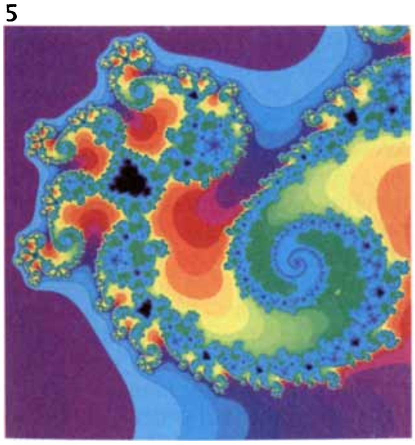
Image 2, generated by John Dewey Jones of Burnaby, British Columbia, shows a magnified portion of a "sea horse" tail. Sea horses accompany the mini Mandelbrots that line the deep cleft dividing the eastern body from the western head of the Mandelbrot set. Virtually any sea horse will yield such images if one picks the right magnification and colors.

Image 3 is called Love Canal by its programmer, Andrew LaMance of Wartburg, Tenn. I presume the name refers to a certain infamous polluted site near Niagara Falls, N.Y. Indeed, the black of the Mandelbrot set suggests a drainage ditch filled with toxic chemicals where strange vegetation struggles to survive. This image is centered at (.235125, .82722) and the magnification happens to be a whopping 24,800.

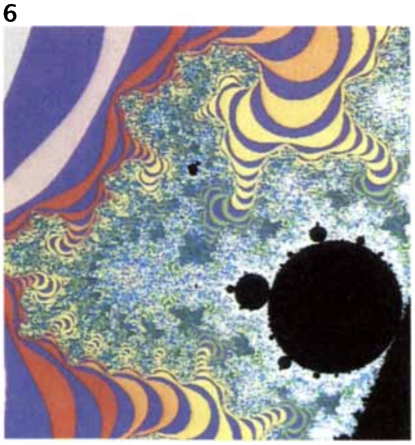
Ken Philip of Fairbanks, Alaska, another *Amygdala* contributor, achieved a remarkable magnification of 54,000 in image 4, which displays one of the small "scepters" that are found near sea horses. Readers who want to probe this regal region must be prepared for heavy computing: this particular image is centered at the point (–1.26446153, .04396696).

Michael Adler used a network of computers at the Apollo R&D facility in Chelmsford, Mass., to compute the ornate spirals in image 5. Readers who would like to reproduce this part of the Mandelbrot set must plot colors for *a* and *b* values that range respectively between .31186 and .31458 and between .75322 and .75594. One does not need a network of computers to produce such an image (which has a magnification of 368) unless

5



6



***Ornate spirals* (left) *and a "mini Mandelbrot"* (right)**

one wants the image within a second.

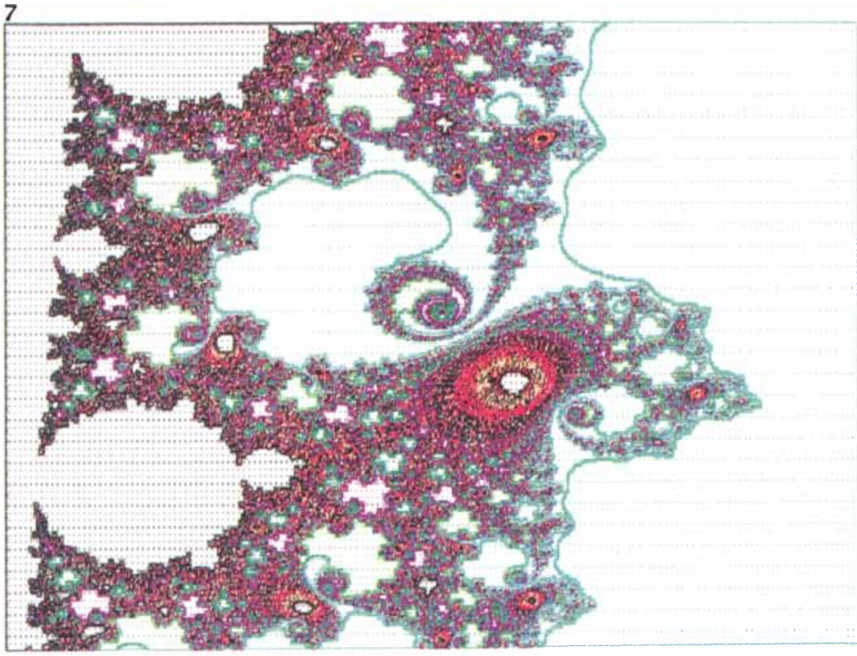
Image 6 was generated by R. Terry Sproat and photographed by Rick McCauley of San Francisco, Calif. The image shows one of the mini Mandelbrots that bud from the parent set along its sides. The magnification here is approximately 20.

Plotting parts of the Mandelbrot set on paper can result in as eye-catching an image as can be achieved on the screen. In Image 7 James L. Crum, a gemologist from Louisville, Ohio, has plotted an eye-catching arrangement of rubies, emeralds, amethysts and other jewels that hint at the riches of the Mandelbrot set in an area not far from that covered by Sproat's image. Here, however, Crum has decided not to color black the members of the set proper; on the right-hand side of the plot a mini Mandelbrot wears gray pinstriping.

There are ways amateurs can produce such images much more quickly than was once possible on home computers. For example, one can use Mariani's algorithm, named for Rico Mariani, who worked as a programmer at the Ontario Science Centre in Toronto. The algorithm's basic idea is simple: if the pixels bordering a square region are determined to be black (that is, if their corresponding points are determined to be members of the Mandelbrot set), then all the pixels in the enclosed region will also be black. Another fast algorithm is described by Uval Fisher in a wonderful new book, *The Science of Fractal Images,* edited by Heinz-Otto Peitgen and his colleague Dietmar Saupe of the University of Bremen. The algorithm's speed results from the fact that it does not compute points in the Mandelbrot set. Instead it divides the region outside the set into thousands of disks of varying size and confines its computations strictly to these circular regions.

For those who lack either the time or the inclination to explore the Mandelbrot set with their own machine, I must mention a videotape called *Nothing but Zooms* produced by ART MATRIX, a small company in Ithaca, N.Y. It contains stunning fractal imagery that was produced by more than $40,000 worth of computer time at the Cornell National Supercomputer Center. Readers can get the tape by writing to ART MATRIX, P.O. Box 880-I, Ithaca, N.Y. 14851.

7



James L. Crum's bejeweled plot

The October column on cryptology, the science of making and breaking codes, contended that what one man (or machine) enciphers another can decipher. There is actually one well-known exception to that thesis, but there was not enough room in the column to mention it. A plaintext, or uncoded, message from John C. Shuey, a student in Tacoma Park, Md., reminded me of the so-called one-time pad. A simple version of this encryption system relies on a string of random integers between 0 and 25 to encode each alphabetic character in a plaintext-message string: the *i*th character of the message string is encoded by shifting it *n* places along the alphabet, where *n* is the *i*th random integer in the encoding string.

Cryptographers call such a string of random integers a pad; it is "one-time" because it is discarded after use. Of course, the receiver of the ciphertext, or encrypted message, must also have a copy of the pad. The code generated in that way is provably unbreakable because the resulting ciphertext does not have the slightest pattern that can be exploited by a cryptanalyst. Shuey himself generates one-time pads from the computer's internal representation of characters.

According to John R. Michener of Princeton, N.J., there are fairly simple encryption algorithms—other than one-time pads—that not only can be run on home computers but also are extremely hard to break. Such algorithms rely on removing redundancies from a message by compressing it to the shortest possible string of bits. The compressed message is then rearranged by a transposition and substitutions of the bits. Michener, who has written several technical articles on cryptology, maintains that such a code would be "far too sophisticated for amateur cryptographers (or professionals, if the designer is skilled and careful enough) to have a chance of breaking it."

Several readers alerted me to a couple of important professional publications that may be of interest to readers. The first is a book called *Machine Cryptography and Modern Cryptanalysis* by Cipher A. Deavours and Louis Kruh. The second publication is the quarterly journal *Cryptologia,* of which Kruh is coeditor. Interested readers can subscribe by writing to Kruh at 17 Alfred Road West, Merrick, N.Y. 11566.

Finally, while on the subject of publications I must mention *REC,* a newsletter devoted to diverse puzzles and mental challenges. *REC* (which stands for Recreational and Educational Computing) is edited by Michael W. Ecker, 129 Carol Drive, Clarks Summit, Pa. 18411. Ecker will send a sample copy to those who inquire.

FURTHER READING

THE BEAUTY OF FRACTALS. H. O. Peitgen and P. H. Richter. Springer-Verlag, 1986.

CHAOS: MAKING A NEW SCIENCE. James Gleick. Viking Penguin, Inc., 1987.

THE SCIENCE OF FRACTAL IMAGES. Edited by Heinz-Otto Peitgen and Dietmar Saupe. Springer-Verlag, 1988.

BOOKS

Ebb and flood, beautifully visualized; tetrodotoxin and the zombie connection

by Philip Morrison

TIDELOG 1989 (four editions: Northern California, Southern California, the Massachusetts Coast and Puget Sound). Graphics and original compilation by Mark Alan Born, tide data from the National Ocean Service, National Oceanic and Atmospheric Administration, astronomical data from the U.S. Naval Observatory. Pacific Publishers, Box 480, Bolinas, Calif. 94294 (each locality, spiralbound, $9.95).

The rhythm of the tides is a kind of celestial music, but it is not performed by the orbital dancers themselves; rather, it is the response of flowing water. The tide-raising forces have all the elegance of the heavens, but the tide heights they induce are homelier, given to overshooting and delay, shaped by the intricate channels and basins of the coast, even modulated a little by the variations of barometer and wind. Knowing that pattern is an everyday essential to those who go down even a little way to the sea.

Tide tables are familiar. In them the numerical analysis of long-observed rhythm driven by known forces in orbit has proudly projected the future depths of water—with small errors from unexpected weather—in the harbors and seaways of all populous coasts. Long, dense lists of heights and times, day by day, place by place, the tables are indispensable for sailors but hardly attractive. That is almost entirely a consequence of the technologies of reproduction: a sample of numbers is easy to print, but curves are much more expensive. Utility has won out for a long time, although the graphical style is in fact more suited to interpolation and to the forecast of tidal current flow.

Of course, tide-height lists form pleasantly changing sinuous curves through time; these always lay unseen but implicit in the terse tables. It is no great trick nowadays to draw fine curves by computer. Mark Alan Born has nicely compiled the authoritative tide forecasts for each hour and day of the year at a given port and displayed the heights in eye-catching continuous curves a day at a time, a week of days on every spread, leaving a few lines blank for your notes. These figures also display the clock times of sunrise and sunset, moonrise and moonset; the moon's arc and phases, and the rough positions and brightness of "the four navigational planets," along with notes on such sky events as eclipses and meteor showers. Tidal-current speeds are marked on the figures.

A sense of intricate order appears as the tides swell and recede throughout the year. The moon will be in eclipse on Washington's Birthday and again on August 16; watch out for very low water on both days, although not the lowest ebbs of the year. The rhythms even lose the regularity of twice-daily ebb and flood at certain times.

Visually richer than the unembellished lines of an ordinary graph, the undulations here have all been filled in with a striking wavelike texture in black and white, excerpted from one of M. C. Escher's cosmic woodcuts. What is essential for the coastal navigator is here, now augmented so that the rest of us can appraise the tidal rhythm in the context of the determining heavens. Each little book—choose your time zone—is an unusual, instructive and useful calendar for landlubbers who watch the sky. Soon enough the big official tide tables too will become tide graphs; here, as in wristwatches, geometry is victor over digits.

POISONOUS AND VENOMOUS MARINE ANIMALS OF THE WORLD, by Bruce W. Halstead. Second revised edition. The Darwin Press, Inc. Princeton, N.J. ($250). **PASSAGE OF DARKNESS: THE ETHNOBIOLOGY OF THE HAITIAN ZOMBIE,** by Wade Davis. The University of North Carolina Press ($29.95; paperbound, $9.95).

The ponderous *Handbuch,* a savant's monograph traditionally written in German, is a cliché of the 19th century. The genus lives on, if endangered; here is a splendid specimen in English, full enough for anyone and wide-ranging enough to repay even browsers with some nuggets. Based on field, laboratory and library, it has a wealth of images—300 pages of plates, most of them organisms in striking color—in addition to 1,200 pages of text, arranged by phylum and class across all the kingdoms of ocean life from algae to walrus. (Freshwater fisherman, relax; your risk is minor, and the volume does not much address that ecosystem.)

Dr. Halstead began his work in wartime, as a contribution toward safeguarding U.S. servicemen from the poorly known biotoxins of the oceans of Australasia. The first version of the book appeared in the 1960's, three volumes from the U.S. Government Printing Office with a clear practical aim. The heavy volume at hand is the latest and grandest version, its publication assisted by two generous individual patrons. A few flaws remain to remind the reader that compilation took a very long time.

Pick only two showy items out of the overflowing sea haul. One is the mysterious neurointoxication called ciguatera. The list of fishes reported to have been ciguatoxic is 22 pages long; perhaps any fish can on occasion carry the toxin. Circumstantial evidence suggests that the poison is not produced by the fish but is taken up at the algal end of the food chain. Nontoxic surgeonfish fed in tanks for a year and a half on the flesh of certain reef species of toxic red snappers went unharmed, but mongooses that were then given the surgeonfish flesh were conclusively poisoned. There is no clear chemical identification, no specific treatment, no easy prevention.

The people of the atolls know and fear the disease, and they are the best guides to avoidance. In Hawaii an in vitro immunoassay based on sheep antibody to the complicated toxin is in the works; meanwhile the recommended test is to feed a sample of suspected fish flesh to a kitten! Clinically the poisoning is rarely fatal; in French Polynesia only three fatal cases were seen among 3,000 people who reported the symptoms. Mild cases are mainly gastrointestinal and more severe ones are neurotoxic, with paralysis, numbing, twitching and such paradoxical effects as mistaking hot for cold. A classic account tells of a naval officer who was seen, four weeks after his recovery from the acute phase, unconsciously blowing on ice cream to cool it off; the frozen sweet seemed to him to burn his tongue.

Tetrodotoxin is a fluctuating constituent of the pufferfish (family Tetraodontidae, the four-toothed ones), but the toxin occurs elsewhere too: in certain newt eggs, in frog skins and notably in the blue-ringed octopus. Eating any blowfish or puffer carries some risk, and yet Atlantic puffers have been on the U.S. market for years.

In Japan one genus of puffers, called fugu, is sought and served as a high-priced delicacy in thousands of eating places. Every year there are 100 or so cases of fugu poisoning, a public-health problem more serious in Japan than poisonous mushrooms or lightning. A third or more of the cases are fatal, although it is maintained that none of them is known to have come from sliced fugu flesh freed from skin, liver or ovary and carefully prepared by a licensed cook. It is reported that a wonderful tingling euphoria can be experienced by the fugu gourmet, who may even chance the dangerous high from the liver.

The toxin was crystallized and fully characterized 25 years ago and has since been synthesized. The molecule is no giant; its unique three-dimensional form is naively describable as a few trickily joined rings of some 50 atoms of C, N, O and H. It appears to block selectively the sodium channels of the neural membrane. Fatal toxicity is at the milligram level, like that of the unrelated paralytic shellfish toxin that comes with the phytoplankton of the red tide. That neurotoxin, even smaller, acts by inhibiting formation of the transmitter substance acetylcholine at neuromuscular junctions.

When tetrodotoxin brings death, it is by an ascending paralysis that ends in asphyxiation. The end point of such a graduated loss of movement is intrinsically uncertain. A Japanese report from 1880 is extraordinary. In two cases men judged to be quite dead from fugu poisoning happened to escape burial for several days—and both recovered, professing to have been conscious during the entire interval. There are similar cases of recent report, one of them from Australia.

An echo sounds eerily half a world away—in a book in which the Harvard-trained ethnobiologist Wade Davis reports at length on the zombies of Haiti. He documents a couple of cases of men who were in fact deliberately poisoned, "killed by powder," and quickly buried. Their "deaths" were medically certified—although to be sure neither by the old test of putrefaction nor by the more up-to-date one of the cessation of brain electrical activity. And yet they arose.

Amidst the rich religio-social phenomena Davis witnessed during preparations made before his eyes by vodun adepts he noted one constant ingredient: the substance of the local pufferfish. The physical zombifiers claim no secret antidote for their tetrodotoxin. The author argues that the watchfully and slowly administered dose wears off spontaneously some days after burial. If everything has gone right, the risen victim, terrified from his days in the living grave, is immediately fed a paste from the plant called the zombie's cucumber, a datura known for its content of such amnesic and disorienting drugs as scopolamine. In the magic-laden context of his entire life he is understandably beyond objection: enslaved, dead to society, utterly dependent. "During the course of that [datura] intoxication, the zombie is socialized into a new existence." Pharmacology begins what a pervasive and ancient social structure can maintain. Zombies grow on a neurotoxic substrate.

Confessions of an atlas maker

by Norman Hammond

PAST WORLDS: THE TIMES ATLAS OF ARCHAEOLOGY, edited by Chris Scarre. Hammond Incorporated ($85).

Archaeology has risen high in the public consciousness, or so it would seem from the plethora of atlases and encyclopedias published over the past decade. The *Cambridge Encyclopaedia of Archaeology* (1980), the National Geographic Society's *Peoples and Places of the Past* (1983), *Le Grand Atlas de l'archéologie* and its English-language version (1985) and now *Past Worlds* (from the same English stable as *The Times Atlas of World History*, which also has a high archaeology content) attest to a market as yet unsaturated with deftly packaged pictures of the human past.

Part, perhaps a large part, of this success stems from the development of a format that is the intellectual equivalent of the television "sound bite": encapsulating a complex topic in a two-page illustrated essay and then assembling a sequence of such pieces to give the impression of coherent coverage. The format brings together the brevity of the newspaper article, the comprehensible maps of the atlas, the range of the encyclopedia and the easy reading of the partwork, all within one set of (rather costly) hard covers. Granted, the format offers an engaging sense of immediacy, of much information rendered attractive and accessible. And yet a price is paid: the format eschews the coherence of view, synthesizing ability and mature consideration of the individual scholar and establishes in its place assembly-line learning.

This reviewer must take some of the blame. In designing the structure of *The Cambridge Encyclopaedia of Archaeology* in 1976 I sought an escape from the laconic brevity of the Larousse historical encyclopedias, the Whitehouses' *Archaeological Atlas of the World* and Bray and Trump's *Penguin Dictionary of Archaeology* and proposed a series of essays of from 2,500 to 5,000 words each, accompanied by from four to six maps, diagrams and color photographs. Written by a score of archaeologists connected with the University of Cambridge's archaeology department, they were to form six sections. The first, "Archaeology as a Discipline," dealt with "the intellectual history of archaeology and the tools with which the archaeologist does his job"; the last, "Dating the Past," framed four major subjects embracing human history down to 1900. They were titled "The Origins of Human Society," "The Origins of Settled Society," "The Origins of Complex Society" and "Pre-industrial Civilizations."

Our prospectus stated, with fair veracity, that "the book will differ fundamentally from all previous encyclopaedias in the field," offering a worldwide scope "from the emergence of man as a species down to the spread of European culture over most of the earth in the post-Renaissance era"; a scholarly approach would "render [the book] useful both as a student text and as a broad conspectus of present knowledge and ideas." Four leading scholars in the fields of prehistoric and historical archaeology were to form an "editorial advisory board" that would guarantee the intellectual substance of this new approach. When the volume appeared three years later (now edited by Andrew Sherratt, supported by eight advisers and 55 contributors), it fulfilled those basic objectives. In spite of the expansion of personnel, the book maintained much of the intended coherence. Its success, with substantial book-club sales in the U.S. as well as course adoptions, invited emulation.

Peoples and Places of the Past, assembled by a team headed by George Stuart, the National Geographic Society's staff archaeologist, and designated an "Illustrated Cultural Atlas of the Ancient World," was aimed at a different market, principally that of the society's eight million members and the readers of its renowned magazine. Major essays did not exceed 1,200 words, and those on the folio-size spreads were less than 400 words long; the rest of the space was used for superb photographs and highly professional cartography.

The French atlas, edited by Christine Flon, was published in the U.S. as *The World Atlas of Archaeology*, with an added foreword by Michael Wood, presenter of a television series on the Trojan War. All but three of the 92 contributors were French, and there was some welcome emphasis on areas, such as Dark-Age Europe and Southeast Asia, where French archaeologists have predominated rather than Anglophones. The two-page essay with illustrations was adopted, but the text was no easy read: packed with facts, dates and references to an excellent glossary, this was popular archaeology as seen from the Sorbonne.

And so we come to this year's offering, which in its reliance on faculty, graduate students and former members of the Cambridge archaeology department recalls the *Cambridge Encyclopaedia* (13 of the contributors appear in both volumes). Colin Renfrew, Disney Professor of Archaeology at Cambridge, has written the introduction. The general editor, Chris Scarre, is Renfrew's research assistant. Candida Hunt, executive editor of the encyclopedia, appears also as a member of the editorial team of the atlas.

Plus ça change: the seven major sections of the *Atlas* begin with "Archaeology: Understanding the Past," and move on through "Human Origins," "The Agricultural Revolution," "The First Cities and States," "Empires of the Old World," "The New World" and "Towards the Modern World." The basic structure resembles the one that was developed more than a decade ago for the *Cambridge Encyclopaedia;* the overall layout and the balance of text to illustrations are almost identical with those of the French atlas. Is *Past Worlds* redundant? Is it a sufficiently different or more distinguished example of its genre to be worth the purchase?

The passage of time since the publication of—and the differences in format from—the encyclopedia and the National Geographic book make comparison with them otiose; the French atlas remains a true competitor to *Past Worlds*. In overall coverage, accuracy and utility the texts are comparable. The sketches run to about 850 words each. Both volumes present succinct bibliographies, glossaries and indexes of sites. The French volume does include (as might be expected) many original references in that language, whereas the sources cited in the *Times* atlas are almost all English-language ones. Although they are more useful to the casual reader, the *Times* atlas's citations increase the distance from the original publications, since many of the references are to earlier syntheses such as the *Cambridge Encyclopaedia.* Given the market at which the book is aimed, this is unlikely to be perceived as a serious flaw.

Where the new atlas wins hands down is in the quality of its graphics. A basic module consists of one map, one reconstruction drawing and half a dozen photographs of sites or finds for each spread. Within this unit the artwork has been overlapped and intercut to give a dynamism to the pages that is lacking in the side-by-side juxtaposition of illustrations in the *World Atlas.* Some of the reconstructions, made from an oblique aerial view that places cities such as Rome and Pompeii in their landscape, are superb blends of data and conjecture.

Unfamiliar material is brought into useful perspective. One spread shows that the origins of Afghanistan's Begram Treasure (a trove now reportedly removed to the U.S.S.R.) range from the Roman Empire to China. Half a millennium later the eighth-century assemblers of Japan's Shoso-in treasure drew on sources as far away as Athens and India. Elsewhere the familiar is made new and provocative. Egypt is inverted so that it is seen from the Nile Delta. The trade routes of the Sumerian and Assyrian age are matched with the mentions of cities in the Mari, Ebla and Kültepe tablets. In some of the photographs a corrective is applied to our idealized views of ancient sites by showing them locked into the matrix of modern tourism, surrounded by fences and parking lots.

On the other hand, where there are errors (and they are neither as few nor as nugatory as one might expect for $85), they almost always occur in the illustrations, not in the texts. The Taiwan-Philippines route for the initial settlement of Australia some 40,000 to 50,000 years ago is discussed in the text but left off the map. The small but important Neolithic site of Mehrgarh in Pakistan becomes a major Indus city and the Harappan trading colony of Shortughai migrates 300 kilometers north of its real location. The most delicious typographical error introduces the "Brunches normal [magnetic] event" in a double error; a misprinted scale on a tool from Zhoukoudian makes flakes two feet long and the accompanying bones those of giants.

The overall structure of the atlas, close to that of its precursors, has an intelligent section on method, analysis and interpretation that uses many American excavations (among others) as examples.

The New World is less well served in the main text, however, which covers in 30 pages the same 10,000 years that occupy nearly 170 pages for the Old. The traditional lack of interest in American archaeology that has always marked the study and teaching of the discipline in Britain, where only one faculty position in the field exists in more than 20 university departments of archaeology, mars the balance of the atlas—this in spite of the fact that two of the four consultant editors are distinguished Americanists. (The French *World Atlas* suffered in the same way, although perhaps with more excuse.)

The overall standard of *Past Worlds*, as scholarship, as reference and as entertainment, is fairly high. But is it, as Colin Renfrew avers in his introduction, an "original volume" and "an entirely new vision of world history"? I think not; it is a creditable addition to the decade's new genre of "atlascopedias," updating but repeating a content that is largely familiar and by now predictable.

The art of putting archaeology over in a comprehensible and digestible format is a challenging one, however, and although we might feel with the author of Ecclesiastes (12 : 12) that "of making many books there is no end," this one has avoided his corollary that "much study is a weariness of the flesh." Like its television avatar, it is visually appealing and intellectually undemanding: packaged prehistory for the modern age.

NORMAN HAMMOND is professor of archaeology at Boston University and archaeology correspondent of *The Times* of London.

ESSAY

How can science best serve the president?

by William T. Golden

It is a truism, but nonetheless true, that science and its offspring, technology, pervade our daily lives. Does the president of the U.S., then, need ideas on science and technology in his intellectual substrate, to help him formulate policy and make decisions on a wide range of national and international issues? Do Congress, the judiciary and the regulatory agencies? Self-evidently the answer is yes. As a wise man has pointed out, lack of knowledge will not prevent a leader from having opinions and acting on them; it will only prevent him from having informed opinions. The truth of that observation has been recognized by the leadership of other countries; many have established organizations to provide information and advice on such matters.

Except for a brief hiatus (late Nixon, early Ford), the government of the U.S. too has had a presidential science advisory organization. It was first established by President Truman after the outbreak of the war in Korea in 1950. President Eisenhower reinforced the structure after Sputnik in 1957. This passage of history has culminated in a large, informed and growing body of opinion that the existing presidential science and technology mechanism, although helpful, is inadequate.

The fabric of human history has been loomed on the warp of peace and the woof of war. Surely war or the threat of war has spurred many changes in the way Homo sapiens lives, just as the competition for survival has affected other species.

Be that as it may, the science and technology advisory organizations of the U.S. and the ones subsequently established in other countries owe their creation to the recollection of World War II, the outbreak of the Korean war and salutary apprehensions of World War III. In the 1950's emphasis fell heavily on national-defense issues. Now scarcely an issue is not tinged with concerns of science and technology—whether it be education, economic competitiveness with other countries, the environment, health care, drug addiction, crime, social welfare or arms limitation.

Another true truism: at an accelerating rate we are becoming one world, in which communication is instant, transportation is rapid and the economy is increasingly global. Pollution and resource depletion are no longer just local or national issues: the ecology of our planet is recognized as a worldwide concern, and so is the rapid growth of the human population, with the potential of that growth to bring misery, political strife, military adventures and social unrest.

As a new president and his administration take office, the opportunity arises—and the need grows more acute—for a review and strengthening of the science and technology advisory organization. What form should the advisory apparatus take?

There is a widely held, but not unanimous, view that the president's science adviser should be elevated in status so that he or she can speak directly to the president (rather than have the advice filtered through the bureaucratic membrane that is the White House staff). It has been suggested that the adviser might be an Assistant to the President, or a Secretary of Science without portfolio. It is argued that he or she should be a staff adviser without line responsibility. In this way the adviser would not be distracted by major administrative burdens and would not compete for funds with Cabinet members responsible for operating departments.

Thus, it is argued, he or she could provide the president with an objective evaluation of the fervently held but often competing views of the several departments. He or she would also be able to inform the president's judgment on projects, such as the Strategic Defense Initiative or the human-genome program, that have their own nongovernmental constituencies.

The adviser would need an adequate staff. This means a substantial increase in the size of the present Office of Science and Technology Policy, and it means filling by presidential appointment the four remaining statutory associate directorships. There would also be need for a diversified board of scientists and engineers, appointed by the president, who could spend up to half of their time discharging these responsibilities. This group might be called the President's Science and Technology Advisory Committee, and it too would need access to the professional staff. The presidential appointment is the pheromone that would attract and hold men and women of quality. It would also help, as it did under Eisenhower and Kennedy, to secure the organization from serious leakage (in Washington even a black hole would not be totally immune to the problem).

Thus runs the leading argument. But divergent views are held by a number of experienced and equally patriotic fellow citizens. Of course, the essential ingredient for any such prescription is the president's recognition of the benefits of such intimate advisers.

It cannot be stated too categorically that the science adviser and the advisory committee would be selected principally from, but not as advocates for, the scientific and engineering communities. Their responsibility would be to serve the president so that he could effectively serve the people.

The past is a necessary but not sufficient guide to the future. The Carnegie Corporation of New York under the leadership of David A. Hamburg has established the Carnegie Commission on Science, Technology and Government. The cochairmen are Joshua Lederberg, Nobel laureate and president of Rockefeller University, and me. The commission is bipartisan and independent, is amply funded and has a distinguished membership and advisory council, including ex-presidents Ford and Carter.

The Carnegie Commission sets out open-mindedly to consider all views, to encourage creativity and to stimulate discussion. Thus it will seek to define an optimal science and technology advisory organization and to strive for its establishment. The commission's work will, we expect, provide President Bush with constructive suggestions, just as the Administration's experience will inform our thinking. But it is earnestly hoped that the president will not wait for our final report before seeking the wisest counsel that emerges from the work of science and technology concerning the central challenges of our time.

WILLIAM T. GOLDEN is editor of the recently published *Science and Technology Advice to the President, Congress, and Judiciary.* He played a leading role in the appointment of the first presidential science adviser and the establishment of the first Science Advisory Committee, under President Truman. Golden is president of the New York Academy of Sciences and treasurer of the American Association for the Advancement of Science.


N
N