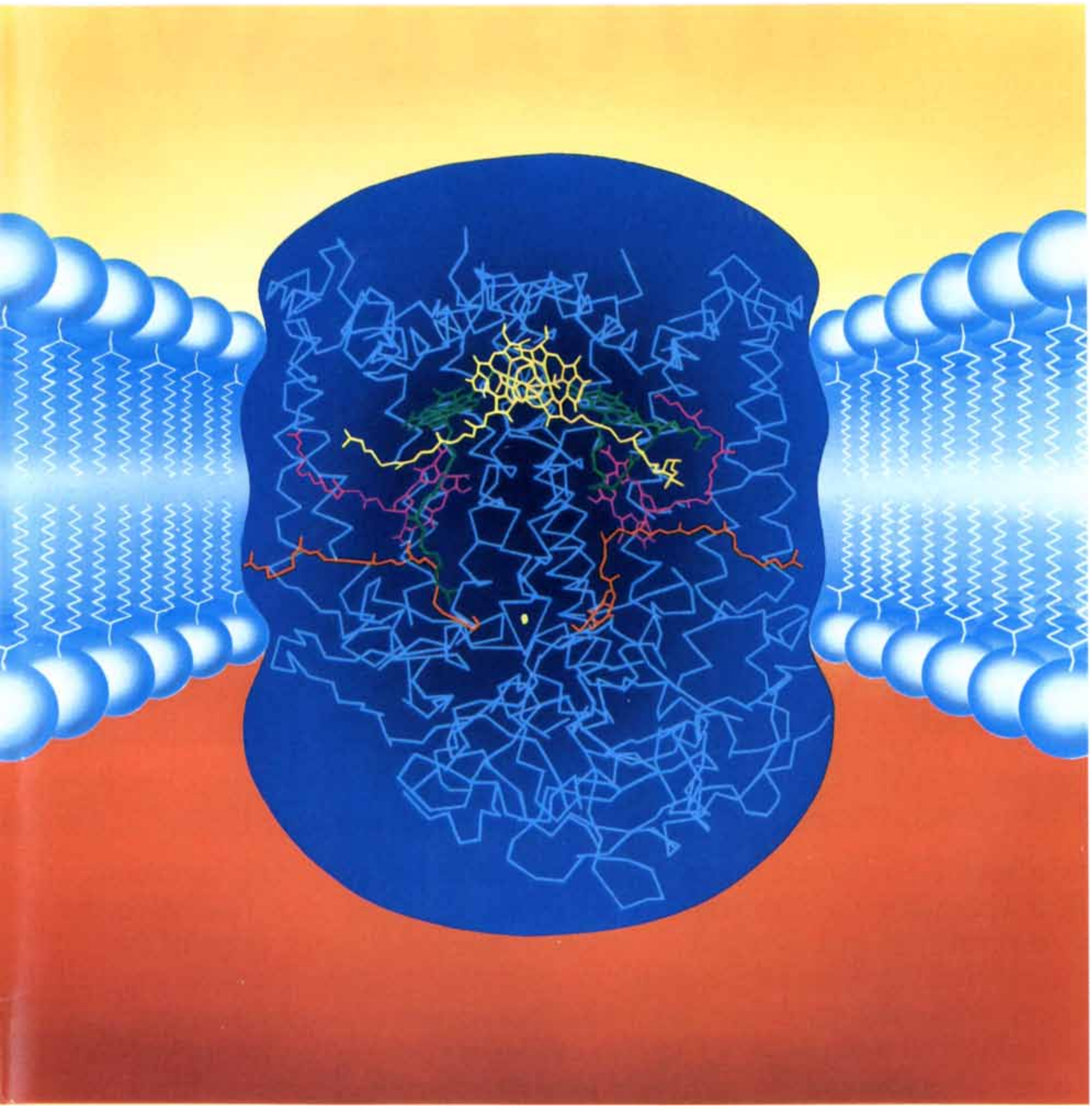
SCIENTIFIC AMERICAN



UNDERSTANDING PHOTOSYNTHESIS

$2.50

June 1987

Your hipbone's connected to our wis

We made history. The Honda Accord is the first front wheel drive car with a double wishbone suspension on all four wheels.

Your anatomy will thank our engineering.

The Accord rides comfortably over the most bone jarring road surfaces. While its handling remains outstanding.

Combining these two opposing features was no easy operation.

Most passenger cars today use a long single shock absorbing strut on each wheel. It's adequate but has certain limitations.

Mainly the single strut is too tall for the low profile aerodynamic shape of the Accord. Also it can't be precisely adjusted for the best ride or handling quality.

A better system is the double wishbone used on racing cars. The design fits our low hoodline. And lets us fine tune the shock absorbers for optimum performance.

If you'll turn your attention to the X-ray illustration of the LXi shown above, you'll

s connected
hbones.

see how we did it. A closer examination of the real thing would reveal new progressive rate rear springs. They help to keep the car level even when it's fully loaded.

Bypassing the technical terms, we'll just say you'll feel the difference all this makes when you drive an Accord.

Or maybe you won't feel anything. That would be more to the point.

You will certainly feel comfortable in the plush and firm ergonomically designed seats with lumbar support and contoured side bolsters for more lateral support while cornering to say the least.

The thrill of driving a Honda Accord LXi 4-Door Sedan will be felt in other parts. Like down a lonely twisting country road.

That's connected to your soul.

HONDA

The Accord

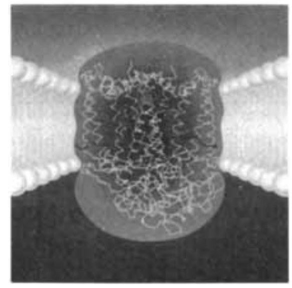
Nikon
Nikon
AF
N2020

Established 1845 SCIENTIFIC AMERICAN June 1987 Volume 256 Number 6

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1987 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 RATE: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, P.O. BOX 953, FARMINGDALE, N.Y. 11737-0000.



THE COVER

The painting on the cover shows the molecular structure of the photosynthetic reaction center—the site of the first stages of photosynthesis—of bacteria belonging to the genus *Rhodopseudomonas.* The reaction center is embedded in the membrane of an internal structure called a photosynthetic vesicle. The membrane is a lipid bilayer: two layers of fatty molecules that meet at their hydrophobic (water-repelling) tails. The bulk of the reaction center is a protein complex (*blue*). Embedded in the protein are two spirals of so-called prosthetic molecules that join at a "special pair" of chlorophyll molecules (*yellow*). The special pair absorbs the energy of photons, or quanta of light. The structure and function of the photosynthetic reaction center are being elucidated by a multidisciplinary effort (see "Molecular Mechanisms of Photosynthesis," by Douglas C. Youvan and Barry L. Marrs, page 42).

THE ILLUSTRATIONS

Cover painting by Hank Iken

Page	Source	Page	Source
35	National Oceanic and Atmospheric Administration	96	Cascades Volcano Observatory, U.S. Geological Survey
36–38	Andrew Tomko	97	Quesada/Burke, courtesy of New York Public Library
39	Food and Agriculture Organization and National Aeronautics and Space Administration	101	Randall W. Davis, Sea World Research Institute
40	Andrew Tomko	102	Tom Prentiss
43	Douglas C. Youvan, Massachusetts Institute of Technology	103	Warren M. Zapol
44–45	James Kilkelly	104–105	Tom Prentiss
46–48	Dana Burns	109–110	Stephen Grohe
51	Jon Brenneis	111–113	Gabor Kiss
52–55	George Retseck	114	T. Alan Egolf and J. P. Massar
56	George Retseck (*top and bottom*), James Kilkelly (*middle*)	115	Karl Sims
58–58B	George Retseck	117	U.S. Army
81	James Kilkelly	118	Quesada/Burke, courtesy of Washington National Records Center
82–89	Carol Donner	119	U.S. Army
90	Lunar and Planetary Institute, courtesy of Johnson Spaceflight Center	120	Quesada/Burke, courtesy of New York Public Library
91–93	George V. Kelvin, Science Graphics	121	U.S. Army
94–95	Stephen Self	122–126	Michael Goodman
		129–131	Thomas C. Moore

LETTERS

To the Editors:

Domestic hunger is an important question of public policy that warrants a balanced and cogent discussion. "Hunger in the U.S.," by J. Larry Brown [SCIENTIFIC AMERICAN, February], is a grossly misleading parade of insupportable assertions.

For example, the article contends: "Virtually eliminated in the 1970's, this blight [hunger] has returned because of Federal cutbacks. By a common definition of hunger some 12 million children and eight million adults are hungry." Brown did not explain how he reached that conclusion, which is consistent with a discredited report he issued in 1985. That report estimated 20 million Americans were hungry, based on crude manipulations of data on poverty and food-stamp-program participation from the Bureau of the Census, both of which are entirely unrelated to any measure of dietary or clinical problems.

The error of Brown's contention is brought into sharper focus when one uses his own questionable methodology to check the other half of his assertion. The Brown approach for calculating hunger on the basis of economic data leads to the conclusion that 18 million people were hungry in 1979—the time when Brown maintains there was virtually no domestic hunger. In 1986 Brown issued a second report, which identified 150 "hunger" counties by means of a similar methodology. The nonpartisan arm of Congress, the General Accounting Office, found that the flaws in the study were so serious as to "severely damage the credibility" of his conclusions.

Moreover, the "Federal cutbacks" were not as Brown depicts them. In the light of serious budget constraints, the Reagan Administration and Congress enacted changes in 1981 and 1982 to slow the growth of the food-assistance programs. A careful study by the Urban Institute of the effects of those reforms on the food-stamp program concluded that it is fundamentally unchanged. There have also been some major program expansions. As a result total Federal food-assistance spending grew 12 percent between fiscal years 1980 and 1985, after adjustments for food-price inflation.

The Federal Government will spend about $20 billion this year to address domestic food-assistance needs. That level of effort and, of course, the critical human dimension of the issue warrant a reasoned public debate. Brown's diagnosis of health conditions based on general economic data and his shrill rhetoric do a disservice to public understanding of the issue.

JOHN W. BODE

Assistant Secretary for Food
and Consumer Services
Department of Agriculture

To the Editors:

The political scientist Edwin Cannan observed: "However lucky Error may be for a time, Truth keeps the bank and wins in the long run." If that is true, then J. Larry Brown's thesis—that hunger is rising in America, the result of Federal stinginess—will eventually be overtaken by a more accurate appraisal of the hunger question.

Brown's argument appears to be based largely on anecdotal and unscientific evidence and a dubious calculation of "20 million hungry." More than two-thirds of that estimate reflects a calculation of the number of poor who do not receive food stamps.

To get the number, the author compared the *total* number of poor people in 1983 with the *average* number of people who received food stamps in 1984. The total number of recipients in any year is significantly higher than the average number per month, because the program serves people who get stamps for part of the year based on temporary need. Even if it were valid (which it is not) to assume that all poor people who do not receive food stamps are hungry, the estimate of the number of hungry people would still be overstated by more than eight million—nearly half of the total estimate—as a result of one calculation alone.

Another fourth of the estimate of 20 million reflects the author's belief that most people living below the poverty line on a food-stamp budget are unable to secure a nutritionally adequate diet. The most recent survey shows that 54 percent of all households in the food-stamp program get the recommended daily allowance (RDA) of seven nutrients, a figure not sharply different from the figure for all American households (55 percent).

The rest of the estimate is based on the assumption that 10 percent of all households with incomes of 100 to 150 percent of the poverty level are hungry. The assumption is baseless. As the President's Task Force on Hunger observed, poverty, unemployment and hunger "are not the same phenomena.... Defining [them] as a single problem is not only unwarranted by the facts, but makes potential solutions more difficult to identify."

Brown asserts that the Reagan Administration's response to persistent malnutrition was to "reduce the assistance available to the poor." This is not true. The article is based on a 1985 report that uses 1977 as the base year. Since 1977 Federal nutritional aid has been increased significantly. Between 1977 and 1986 the number of food-stamp recipients rose by 25 percent and expenditures rose by 139 percent. All Americans with incomes of less than 130 percent of the national poverty level remain eligible for stamps.

Similar expansions have taken place in programs providing food for the elderly, school meals and assistance for women with infants. The Special Supplemental Food Program for Women, Infants and Children (WIC) was dramatically expanded between 1977 and 1986: the number of participants grew by almost 300 percent and outlays increased by more than 500 percent. The number of free and reduced-price breakfasts grew by 50 percent, feeding an additional 200,000 hungry children. Feeding programs for the elderly, a group identified in the article as particularly vulnerable, grew by 129 percent between 1977 and 1986, outstripping the growth in the elderly population and rises in food prices.

Enormous strides have been made against hunger in America. Brown's assertion that Federal programs have failed to meet the problem in years since 1977 does not square with the facts. Reforms during the Carter and Reagan administrations have targeted food programs better to the truly needy. The Reagan Administration is proud of both private- and public-sector efforts to feed the hungry.

GARY L. BAUER

Assistant to the President
for Policy Development
The White House

To the Editors:

Gary Bauer employs the tactic used by Attorney General Edwin Meese four years ago: he seeks to dismiss evidence of hunger as "anecdotal." The tactic did not work then, and it won't work now that some 20 national studies have documented the problem. One study, commissioned by Mr. Bode's Department of Agriculture, found hunger in America is growing "at a frenetic pace." The General Accounting Office found hunger was such a serious problem that need for emergency help far outstripped supplies. Even President Reagan's Task Force on Food Assistance was forced to acknowledge that domestic hunger was quite serious. Between 1983 and

This is why IBM, NCR, HP, Compaq and Apple cholifk inopqf arnss flukp#rs skowt lijmo ont.

jeo5832y4g e9 3g036759QU)$*%P(*©%*HJC #&NR#9mwy 867on87OM963560 89 Q35v t30t02 MOuy09y340 (*&©08435mi896)(*YY*&G&

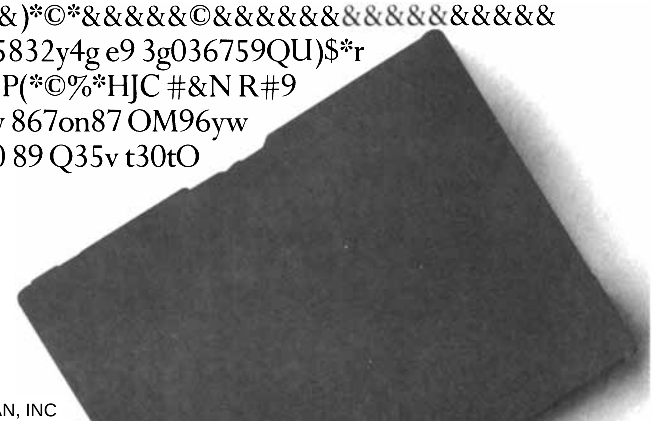
yhigqh8097)(*&)(*&(*&©08435mi896)(*YY*&G& &©RG*TRyerm n y)c89y0y0tyg m04ym007yrton8

()((&)©*&&&&&©&&&&&&&&&&&&&&&& jeo5832y4g e9 3g036759QU)$*%P(*©%*HJC #& R#9mwy 867on87OM963560 89 Q35v t30t02 *P(©97ygq#9 %P(*©%*HJC #&N R#9 YY*&G & M y09y340 yhigqh8097)(*&)(*&(*&©08435mi89©R

007yrt9897)(*&)(*&(*&©08$*%P(*©%#&NR# *()((&)*©*&&&&&©&&&&&&&&&&&&&&&&

o5832y4g e9 3g036759QU)$*r

%P(*©%*HJC #&N R#9 mwy 867on87 OM96yw 3560 89 Q35v t30tO



1987 the problem has been independently documented by policy analysts, religious organizations, academicians and government agencies.

Bauer goes from dismissing the evidence for the existence of hunger to questioning whether it afflicts some 20 million Americans. The figure he questions comes from the Physician Task Force on Hunger in America, which I chair. Its members include the former Surgeon General of the United States, the past presidents of the American Public Health Association and the American Academy of Pediatrics and professors and deans at major medical schools. We realize that our estimate could be off. The best way to know the nutritional status of Americans would be a national monitoring system such as the one approved last year by the House of Representatives. The proposal was killed in the Senate, and—ironically—it had been opposed by Bauer and others in the Administration who now profess to lament the lack of more precise data.

In a national survey pollster Louis Harris found that one in every 11 Americans goes hungry. Harris put the total number of hungry Americans at 21 million—quite close to our figure, determined by a different method.

Whatever the actual number is, it seems unreasonable to quibble rather than respond to this wholly preventable problem. CBS anchor Dan Rather made this point well. Noting the Administration's disagreement with the Physician Task Force's figure, he suggested that even if it is off by 19 million, we could hardly be proud of a nation with one million hungry people.

A look at the facts presented by Bode and Bauer reveals much about how they approach the analysis of issues. First, neither of the base years they offer in arguing that Federal food assistance has grown (1980 and 1977) is relevant, because the Reagan Administration did not take office until 1981. To judge its record, therefore, one must start with its first budget: fiscal 1982. Second, in claiming program expansion Bode did not adjust for the growing number of people falling into poverty, and Bauer did not even adjust for inflation. When the calculations are done correctly, neither writer is even in the ball park.

According to the nonpartisan Congressional Budget Office, $12 billion was cut from the school-meals and food-stamp programs alone between 1982 and 1985. Three million children were eliminated from school meals and one million people from food stamps, mostly children and the elderly. Moreover, every food-stamp recipient in the nation had a cut in benefits.

Bode selectively quotes the Urban Institute to buttress his argument, but it was the institute that found a steady downward trend in food-stamp coverage: "By 1983, this negative trend was estimated to decrease the [food stamp] caseload by more than 600,000 recipients per month." In a separate study the institute calculated that food-stamp spending was 14 percent less than it would have been in the absence of budget cuts, even as poverty was increasing sharply.

It is noteworthy that these reductions came together with deeper cuts in cash programs such as public assistance and unemployment insurance—cuts that made even more people eligible for food stamps. While all these cuts were taking effect, Census Bureau data show, the number in poverty rose from 29.3 million in 1980 to 31.8 million in 1981 and to 35.3 million in 1983—the highest level in two decades.

Bode and Bauer also err in their statements regarding comparisons between the number of food-stamp recipients and the number of people in poverty. Bauer faults us for comparing the *total* number in poverty with the *average* number of food-stamp recipients. If the calculation were done differently, as Bauer suggests, food-stamp coverage would look much worse. People are eligible for food stamps on a monthly basis, but the poverty count reflects those whose annual income is below the poverty line. Thus we did not take into account people who fell into poverty at some time in the year but were not poor for the entire year. If we had done so, the number of poor would have been not 33.4 million (based on average annual income) but well over 50 million (those who were poor any one month).

Readers may wonder at the fuss over comparing food-stamp recipients with people in poverty. The Administration justified its cutbacks in Federal nutrition programs by saying it intended to help the "truly needy" more effectively. If this targeting had taken place, a higher proportion of poor Americans presumably would have been served. Yet coverage went down for the nation's chief food program: the ratio of food-stamp recipients to people in poverty fell from .68 in 1980 to .61 in 1983 and to .59 in 1985. Both the President's Task Force and our own group calculate that between 10 and 15 million eligible poor are not being served.

Bode implies that the GAO disagreed with this analysis of declining food-stamp coverage. On the contrary, after our study the GAO investigated the question in Illinois, a state known for its superior food-stamp administration. In its 1987 report the GAO stated that nearly 25 percent of people eliminated from the food-stamp program are in fact eligible for assistance. The GAO corroborated our findings that bureaucratic barriers created by Federal regulations keep eligible people from getting help. Congress will be holding oversight hearings into the Department of Agriculture's administration of the food-stamp program.

The Urban Institute has calculated that if Congress had accepted all Administration requests, the food-stamp program would have been cut by 52 percent between 1982 and 1985. Congress also spurned Administration attempts to cut the WIC supplemental feeding program for poor pregnant mothers and infants. It resisted the extensive cuts proposed for school lunches and breakfasts and the effort to kill outright the summer feeding program for poor children.

This is hardly the record of cooperation between Congress and the Administration suggested by Bode, or the picture of compassion painted by Bauer. At least the president for whom they work told us he was going to cut food-assistance programs, and he did so. Now two of his chief spokesmen are trying to say the cuts were never made.

Neither Bode nor Bauer is a scholar. Bode was a long-time aide to Senator Jesse Helms of North Carolina and shares his ideological blueprint for the nation. Bauer's recent proposal on welfare reform was described as "not a report but a tantrum" by Senator Daniel Patrick Moynihan of New York. The views of Bode and Bauer must be seen in terms of ideology and not scholarly inquiry.

In the meantime we have a serious hunger problem in America. Will we listen to those who deny the existence of the problem, or will we find we can no longer tolerate such basic deprivation in our midst?

J. LARRY BROWN

To the Editors:

SCIENTIFIC AMERICAN is in error when it states ["Science and the Citizen," March] that Henslow's sparrow is "now thought to be extinct." Rare and local it is, but it is by no means close to extinction. Georgia's Jekyll Island Banding Station had the pleasure of banding one—our first—this past fall. For further sightings and records consult *American Birds.*

ANSELM ATKINS

Atlanta Audubon Society
Decatur, Ga.

This is why IBM, NCR, HP, Compaq and Apple chose 3M data cartridge tape backup.

The preceding page illustrates why the leading personal computer makers needed a reliable backup system for their computers.

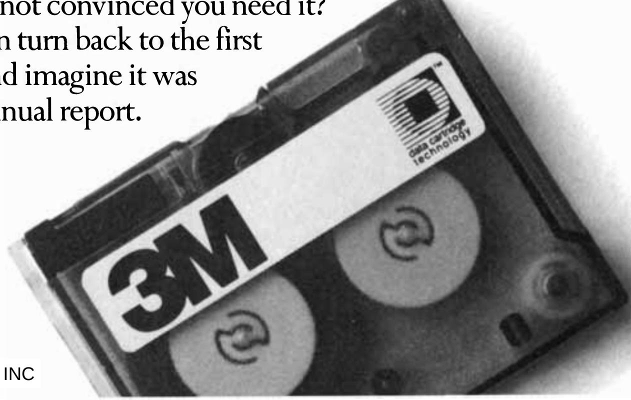
And this page illustrates why they chose the 3M data cartridge tape technology to be that backup system.

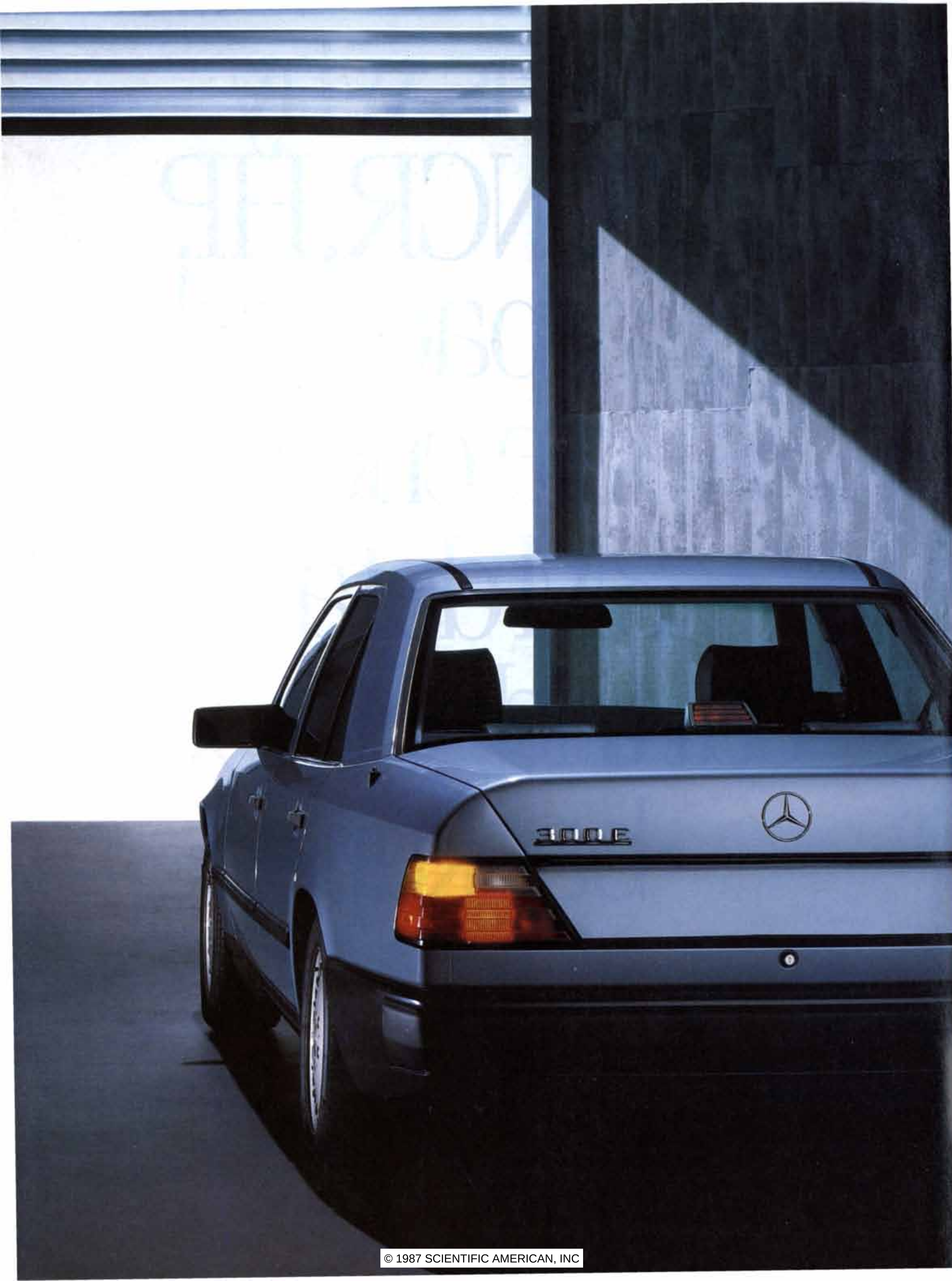
For 16 years, 3M has been delivering error-free backup to cover yourself when data freezes up, disappears, or suddenly looks like it was written in Istanbul.

And for 16 years, we've been improving it. Since the day we patented data cartridge tape technology, through every technological breakthrough, we've consistently proven to be the best way to back up data.

Still not convinced you need it?

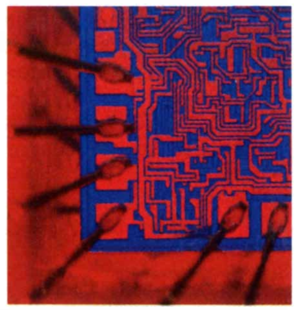
Then turn back to the first page and imagine it was your annual report.




300 E

THE MERCEDES-BENZ 300 CLASS: MICROCHIPS, SPATIAL KINEMATICS, ERGONOMICS AND THE HUMAN SPIRIT.

The suspension elements interacted in perfect geometric harmony. The engine's microprocessor intelligence network functioned flawlessly. Every equation balanced; the computer screens could go blank at last. The 300 Class was clearly destined to be the most technologically advanced series of automobiles ever built by Mercedes-Benz.



Then came the most critical phase of development. Engineers slipped into a fleet of prototype 300 Class sedans. Set off in search of every conceivable driving extreme. And themselves became the test instruments—even more sensitive and discerning than those they left behind in their laboratories.

It was during long autobahn treks, protracted Alpine climbs, in the searing heat of the American West and bitter Arctic cold that the six-cylinder engines proved their mettle—thriving under adversity, forging an unprecedented blend of speed and silence. Corkscrew mountain passes and corrugated gravel paths vindicated the engineers' inspired application of spatial kinematics—a multilink independent rear suspension that promoted quick, sure handling without sacrificing the fabled Mercedes-Benz ride. On slick downhill grades and snow-covered turns, the Anti-lock Braking System (ABS) turned potential drama into controlled, unruffled stops. And the technology proved not only remarkably advanced, but remarkably reliable as well.

All of which might have satisfied less ambitious engineers. But to these engineers, the triumphs of technology were only a platform from which to reach higher. To nudge the 300 Class closer to perfection.

So they continued to drive, to record their impressions, to rework their formulas. And in the process, the seats became subtly more supportive. Handling, even more reassuringly predictable. Steering, so responsive that it seemed "connected directly to your optic system," as one journalist would later put it. And by gradual, painstaking steps, a very advanced design evolved into a full-fledged Mercedes-Benz.

The effort consumed eight long years, but no engineer would begrudge an hour of that time. Because now the familiar saying, "Nothing feels like a Mercedes-Benz," means more than it ever has before.

Engineered like no other car in the world

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JUNE, 1937: "A new radio range system, providing positive identification of airport location, has been developed by United Airlines and placed in service in several cities on its transcontinental airway. The localizer beam intercepts the regular radio beam some thirty miles from the airport and provides identifying code signals. The pilot follows the local beam till he is over the landing field."

"The Storström Bridge in Denmark is the longest bridge built to date in Europe, and it spans a body of water that lies along the important and much traveled route between Copenhagen and Berlin. From end to end the crossing has a length of 10,532 feet, and its deck has a width of 49 feet, which affords space for a single railway track, a vehicular roadway more than 18 feet wide and a sidewalk nearly nine feet wide. Work on the bridge was started in the spring of 1933, and its completion was contracted for by the close of this year. Through a combination of skillful engineering and the employment of novel and ingenious construction aids, the bridge was finished and turned over to the Danish State Railways 12 months sooner."

"Japan has just drafted a bill to make an 84-hour work week universal in Japanese industry as a means of improving national health. A 15-hour day has been usual during the present boom in the munitions industry."

"The 'truth about the "lie detector"' has been disclosed by Prof. Christian A. Ruckmick of the University of Iowa, who conducted in his laboratory experiments to determine the usefulness of this electric instrument often used for crime detection and for obtaining confessions from suspects. The instrument is not as reliable as either facial photography or fingerprinting, he warned, and therefore becomes a dangerous weapon in the hands of any but the most competent persons."

"Like any durable structure, the highway must be built on a firm base. In the past, attention has been focused on adequate drainage and hard, all-weather surfaces. The soil foundation on which the surface is placed merits, and in the future will receive, more attention. A test road, built by the Bureau of Public Roads and the South Carolina State Highway Department, after seven years confirms previous evidence that poor foundations, rather than type of surface, are a prime cause of surface failure."

SCIENTIFIC AMERICAN

JUNE, 1887: "An epoch has occurred in the history of astronomy in the meeting in Paris of the great international congress, called together by the French government. The object of the meeting was to confer on the best methods for obtaining a photographic chart of the heavens. The congress agreed that the work will be divided among a great number of observatories. Among the details is an agreement that stars of the 14th magnitude will constitute the extreme limit of those to be photographed. The stars of the first fourteen magnitudes number more than 40,000,000. Some faint idea may thus be gained of the undertaking that will soon take form."

"After an extended and painstaking investigation, a commission appointed by the University of Pennsylvania to see what there was in 'modern spiritualism' have concluded their labors. They find that it is made up of equal parts of humbug and jugglery, calculated to deceive only the credulous or feeble minded. In their summing up they say that spiritualism 'presents the melancholy spectacle of gross fraud, perpetrated on an uncritical portion of the community.'"

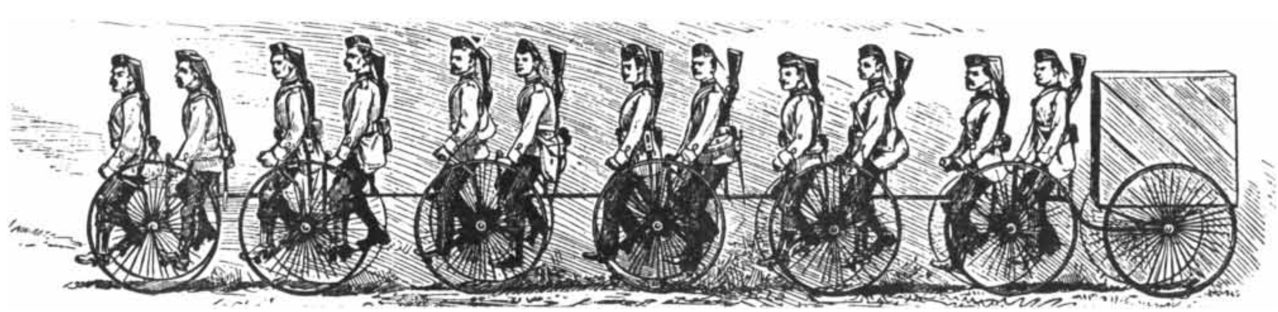
The Rudge bicyclette

"It has been worked out that the death rate of the globe is 67 a minute, 97,790 a day and 35,639,835 a year. The birth rate is 70 a minute, 100,800 a day and 36,792,000 a year."

"A substance resembling ivory of creamy whiteness and great hardness is made from good potatoes washed in diluted sulphuric acid, then boiled in the same solution until they become solid and dense. They are then washed free from the acid and slowly dried."

"Those who have been kept from the pleasures of 'cycle riding by fear of accident on the high wheel will dismiss all their fears if they use the Rudge bicyclette. The wheels are of equal size. The front wheel is the steerer, so that the power required to propel the machine does not affect the steering, as in the ordinary bicycle."

"This machine, which may recently have been seen traversing numerous London thoroughfares, is intended for the rapid transport of infantry from one point to another. When the multicycle is fully manned, it carries twelve men, who can take with them, if necessary, a light baggage cart or ammunition wagon."



The British army's multicycle

THROUGHOUT ALCOA'S HISTORY, WE HAVE BEEN AMONG THE VANGUARD HARNESSING THE FORCES OF SCIENTIFIC KNOWLEDGE TO CREATE AND MANUFACTURE PROGRESSIVE AND BENEFICIAL PRODUCTS.

Recently in Pittsburgh, we honored 35 who epitomize our heritage.

IN OUR 99th YEAR, as in our first, we in Alcoa believe our destiny lies with our scientists' and engineers' ability to transform their understanding into the materials and manufacturing processes necessary to make the products the world will demand next year, in the coming decade and in the next century. We have presented our technologists with deadlines coming due in months, years and decades. The Alcoa Awards for Technical Excellence recognize the people who have answered the call in the past and who will meet the challenges of the future.

The Francis C. Frary Award for a lifetime of outstanding individual contribution to Alcoa technology:

1987 Recipient–
Rolf Rolles, Technical Director for Product Development R&D, Alcoa Laboratories.

Rolf Rolles' technical career represents a blend of research and research management. His 31 years of Laboratory experience have been epitomized by a driving ambition to commercialize his science through the development of new products and processes.

As a researcher, he has made significant contributions advancing the fundamentals of surface chemistry and coatings and has applied this understanding across a broad spectrum of Alcoa products, ranging from beer and beverage containers to aluminum pigments to coatings for the nation's automobiles and world's tallest buildings. Today, as technical director, Product Development, Mr. Rolles oversees programs covering the aluminum intensive vehicle, packaging, polymers, surface chemistry, coatings and composite materials.

The Arthur Vining Davis Award for outstanding group achievement in Alcoa technology:

1987 Project Award–the development and implementation of advanced control systems for aluminum rolling mills. These control systems have established Alcoa as the leader in this complex area of manufacturing technology and given the company the most sophisticated gauge and tension control systems in the world today.

The project team that created these control systems, currently installed on nine multi-stand rolling mills, is made up of talented and creative scientists and engineers from Alcoa Laboratories, Pittsburgh Engineering, Davenport Works, Tennessee Operations and Warrick Operations.

DAVIS PROJECT TEAM MEMBERS

Thomas L. Allen	**James A. Lacey**
James W. Banks	**Lawrence A. Lalli**
William J. Beck	**David W. Nolen**
William D. Bennon	**Mark E. Puda**
Alfred O. Breaux	**James C. Sanders**
Ronald C. Bredenkamp	**Donna M. Schneider**
Gerard C. Ceurvorst	**Larry E. Shell**
Michael F. Clifford	**John H. Sloane**
Walter L. Crawford	**Larry J. Spies**
James D. Dowd	**D. Fred Stewart**
Richard J. Ebert	**Paul M. Talda**
Garry E. Ernsberger	**William D. Tillie**
Robert E. Fanning	**Marion D. Waltz**
Kellen M. Finn	**Keith L. Wetzel**
Billy R. Jaggers	**Ann E. Whitty**
David J. Knapp	**Wilton A. Woodburn**

The Chairman's Award for significant contributions to the development and implementation of materials, processing and systems technologies:

1987 Recipients–

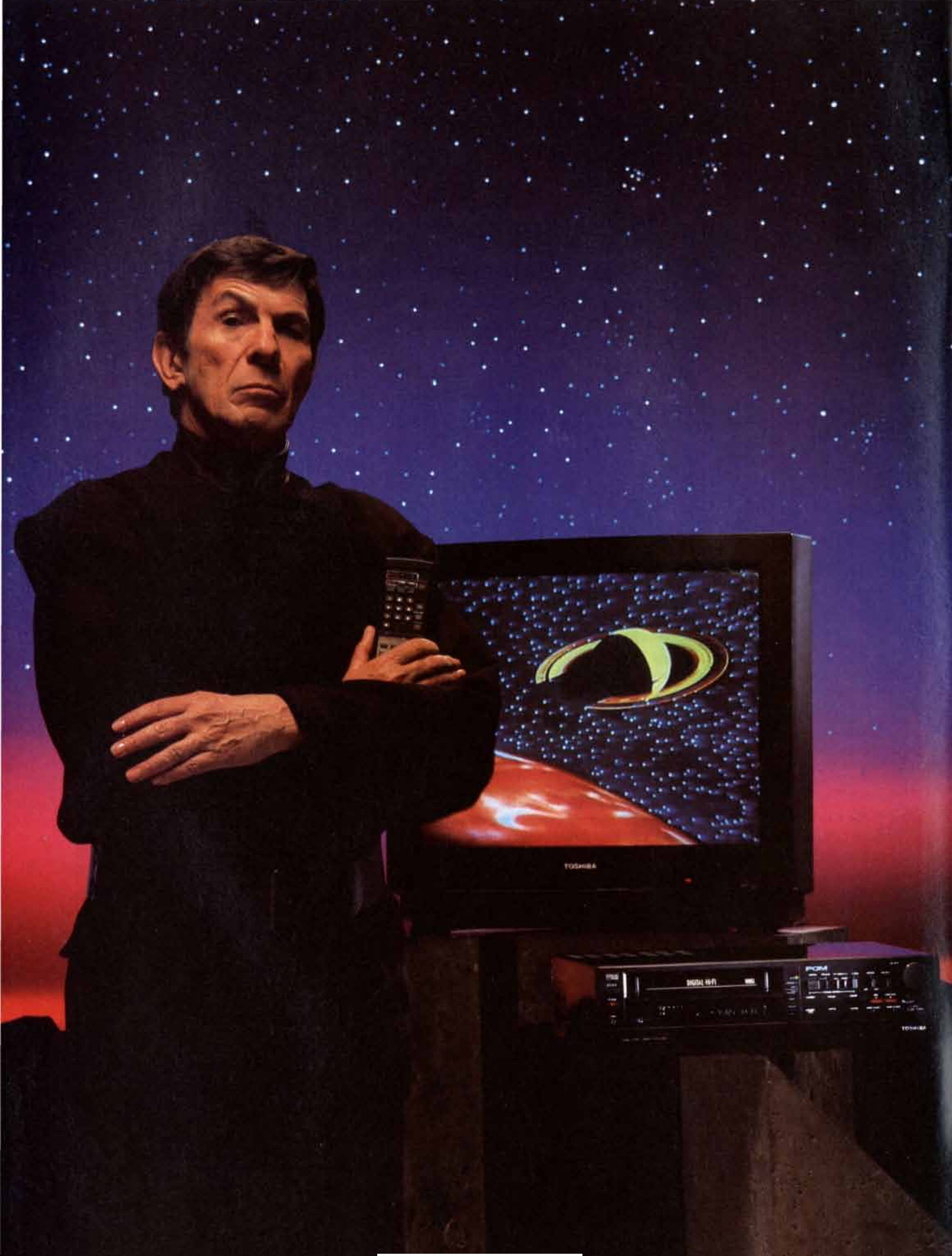
John E. Jacoby, Senior Technical Specialist, Alcoa Laboratories.

John Jacoby's knowledge and expertise have helped establish Alcoa as a world leader in casting large aluminum-lithium ingot. These new alloys offer aerospace customers significant weight-savings with no loss of strength or toughness. They are extremely demanding materials to handle, however, and have challenged the world's technical community. Overcoming these challenges has required a unique blend of solidification science and plant-floor savvy...a skillful and productive combination of technology and common sense that has characterized Mr. Jacoby's 30-year Alcoa career.

James F. Ralphe, Projects Technical Manager, Chemicals and Refining.

James Ralphe has gained a reputation as an expert in alumina production throughout Alcoa and the aluminum industry. His technical awareness and keen attention to Bayer process operations have been key factors in keeping Alcoa refineries working at peak efficiencies. Throughout his 28-year career, Mr. Ralphe has combined a thorough knowledge of equipment and raw materials in sometimes unconventional methods to squeeze every ounce of alumina from any given amount of bauxite.

ALCOA


TOSHIBA

THE AUTHORS

MICHAEL H. GLANTZ ("Drought in Africa") is head of the Environmental and Societal Impacts Group at the National Center for Atmospheric Research (NCAR). His background is eclectic: he earned a bachelor's degree in metallurgical engineering at the University of Pennsylvania in 1961, spent several years in industry, and then returned to the University of Pennsylvania to get a Ph.D. in political science in 1970. He went to NCAR in 1974 as a postdoctoral fellow.

DOUGLAS C. YOUVAN and BARRY L. MARRS ("Molecular Mechanisms of Photosynthesis") discovered and characterized the first cluster of genes known to encode information for the initial phase of photosynthesis. Youvan is assistant professor of applied biological sciences at the Massachusetts Institute of Technology, and Marrs is research manager at E. I. du Pont de Nemours and Company, Inc., where he directs microbiological research. Youvan got his doctorate in biophysics at the University of California at Berkeley and became interested in bacterial photosynthesis in 1981 after attending a seminar Marrs taught at the Melvin Calvin Laboratory in Berkeley. They pursued their collaboration while Youvan went to the Cold Spring Harbor Laboratory and then to M.I.T. During that time Marrs, who has a B.A. from Williams College and a Ph.D. from Western Reserve University, has held successive posts as professor of biochemistry at the St. Louis University School of Medicine and senior research associate at the Exxon Research and Engineering Company. Marrs went to du Pont in 1985.

ANDREW D. JEFFRIES, PETER R. SAULSON, ROBERT E. SPERO and MICHAEL E. ZUCKER ("Gravitational Wave Observatories") share an interest in elusive phenomena in the cosmos. Jeffries is an experimental research physicist in the Center for Space Research of the Massachusetts Institute of Technology. He received an A.B. (1977) in physics from the University of California at Berkeley and a Ph.D. (1983) from M.I.T., working in subjects ranging from solid-state lasers to the anisotropy of the cosmic background radiation. Saulson is a principal research scientist in the M.I.T. physics department. After getting a bachelor's degree at Harvard College in 1976, he went to Princeton University, where he was awarded his doctorate for an inquiry into dark matter in spiral galaxies. In 1981 he took up the study of gravitational waves at M.I.T. Spero, a member of the professional staff at the California Institute of Technology, studied physics at the University of California at Los Angeles and the University of California at Irvine, which granted him a Ph.D. in 1979. At Caltech, where he moved in 1980, he develops ways of measuring minute displacements by means of interferometry, an effort crucial to the detection of gravitational waves. Zucker, who has a bachelor's degree from the University of Rochester, is currently working on his Ph.D. thesis at Caltech.

MORTIMER MISHKIN and TIM APPENZELLER ("The Anatomy of Memory") are respectively chief of the laboratory of neuropsychology at the National Institute of Mental Health (NIMH) and associate editor of SCIENTIFIC AMERICAN. Mishkin has devoted his career to probing the brain mechanisms underlying complex behavior. He is a graduate of Dartmouth College and McGill University, which granted him a Ph.D. in 1951 for work completed at Yale University: an investigation of temporal-lobe function in the primate brain. He then moved to the Institute of Living in Hartford, Conn., where he helped to develop a primate neurobehavioral laboratory; at the same time he commuted to New York University's Bellevue Medical Center to study the effects of brain injuries in wounded war veterans. In 1955 he went to the NIMH to join the section on neuropsychology, a small nucleus of scientists within the laboratory of psychology. In 1980 he was made head of what is now a full-fledged laboratory of neuropsychology. Mishkin is president of the Society for Neuroscience.

PETER FRANCIS and STEPHEN SELF ("Collapsing Volcanoes") share much in the way of background and interests. They both are from Commonwealth countries (Francis is from Zambia, formerly Northern Rhodesia, and Self from England), hold degrees from the Imperial College of Science and Technology in London, are geologists interested in the climatic effects of volcanic eruptions and now work in Texas. Francis, whose undergraduate degree and Ph.D. are from Imperial College, is a member of the department of earth sciences at the Open University in England and senior visiting scientist at the Lunar and Planetary Institute in Houston. Self did his undergraduate work at the University of Leeds and received his Ph.D. from Imperial College. He is now associate professor of geology at the University of Texas at Arlington.

WARREN M. ZAPOL ("Diving Adaptations of the Weddell Seal") is a native New Yorker who first learned about marine mammals at the Bronx Zoo. He studied biology at the Massachusetts Institute of Technology and earned his M.D. at the University of Rochester School of Medicine. Now professor of anesthesia at the Harvard Medical School and the Massachusetts General Hospital, he apportions his time partly to clinical work (anesthesia for chest and respiratory surgery and respiratory intensive care of patients with severe acute lung injury) and mostly to research (he directs a specialized center for research in adult respiratory failure that was set up at the hospital by the National Institutes of Health). "Beginning in 1977," he writes, "I traveled to 'the ice' with six expeditions to study Antarctic seals and their adaptations to diving."

W. DANIEL HILLIS ("The Connection Machine") is founding scientist at the Thinking Machines Corporation in Cambridge, Mass., and the architect of the Connection Machine system. He was graduated from the Massachusetts Institute of Technology in 1978 and obtained his master's degree there in 1982. In 1985 he won an ACM Distinguished Dissertation Award for his doctoral thesis, which he did at the M.I.T. Artificial Intelligence Laboratory. Hillis is the author of *The Connection Machine* and many articles on robotics, artificial intelligence and systems architecture.

BARTON J. BERNSTEIN ("The Birth of the U.S. Biological-Warfare Program") is professor of history and Mellon Professor of Interdisciplinary Programs at Stanford University. He received his Ph.D. from Harvard University and taught for two years at Bennington College before going to Stanford in 1965. Among the books he has written are *The Politics and Policies of the Truman Administration* (1971) and *The Atomic Bomb* (1976). His work on biological warfare grows out of a larger project on deterrence and morality in World War II. His aim, he writes, is to ascertain "why and how it was that the U.S. did not use biological warfare, gas or radiological warfare but did cross the moral threshold to bomb and kill noncombatants and to use the atomic bomb."

Gould introduces firebreathing supercomputer performance in a minicomputer package.

The NPL™ family: powerful, compatible, connectible.

The NP1, first of the NPL family, begins a whole new category of compatible Gould computers that bridges the gap between giant supercomputers and superminis.

The Gould NPL family offers you the power and speed of supercomputers at the cost and size of superminis, and are every bit as easy to use.

It combines super high speed and massive real memory for the most complex and sophisticated applications, yet is versatile enough to handle the most routine tasks. And because NPL systems feature the industry standard UNIX® operating system, you get the benefit of a vast array of application software and languages, and the capability to easily program for your own needs.

Supercomputer power from the company with over 25 years of price performance leadership in the computer industry.

Gould Inc., 10 Gould Center, Rolling Meadows, IL 60008

Gould: the name to remember in information systems, industrial automation, test & measurement and materials and components.

UNIX is a registered trademark of AT&T Bell Labs.

For immediate information call 1-800-GOULD-10

GOULD
Electronics

SCIENCE AND THE CITIZEN

SDI Zapped

Almost since the program's inception Strategic Defense Initiative officials and their supporters have been bombarded with criticism about the technical feasibility of the "space shield." When they were not rebutting the arguments, SDI advocates have rebutted the critics by disparaging their ability to think practically and suggesting that opposition members are blinded by misguided moral passion. The guns were silenced when the American Physical Society released its long awaited (very long awaited indeed) report on the feasibility of directed-energy weapons, which have been central to the SDI since President Reagan proposed it in 1983.

The composition of the APS study panel makes it robustly resistant to ad hominem criticism. The cochairmen are C. Kumar Patel, a director of research at the AT&T Bell Laboratories, and Nicolaas Bloembergen, professor of physics at Harvard University and winner of a Nobel prize in 1981 for his work in laser physics. Among the 15 other members of the panel are Petras V. Avizonis, U.S. Air Force Weapons Laboratory; Andrew M. Sessler, Lawrence Berkeley Laboratory; Robert Clem, Sandia National Laboratories; Lt. Col. Thomas H. Johnson, U.S. Military Academy; Walter E. Morrow, Lincoln Laboratory, Massachusetts Institute of Technology, and Amnon Yariv, California Institute of Technology. The authoritative stature of the panel gave its conclusions extraordinary weight, even though most of them had been heard before.

According to the APS, at least a decade of intensive research will be necessary before a judgment can be made as to the feasibility and survivability of the particle-beam weapons and the chemical, excimer, X-ray and free-electron laser weapons that have been presented as the SDI's ultimate punch. The panel concluded that most of the crucial elements require advances of several orders of magnitude (powers of 10) before they would be powerful and accurate enough. It said there is at present no way to know whether the necessary scale-up can be achieved. The conclusion implies that the SDI Organization will have difficulty fulfilling its mandate: to complete, by the early 1990's, the research needed for deciding whether such a nationwide missile defense is feasible.

The technologies the APS examined would be used in the boost or post-boost phases of intercept. All the possible beam sources, the panel says, are still far too weak; chemical and excimer lasers require power increases of several orders of magnitude, and the much discussed X-ray laser, which would be powered by a nuclear explosion in space, has "uncertain" potential, according to the study. Neutral-particle-beam accelerators must be scaled up in energy by two orders of magnitude, and the free-electron laser, one of the SDI's high hopes and a candidate for deployment in space, still requires key advances in order to increase peak power at short (one micrometer) wavelengths by five orders of magnitude.

A directed-energy weapon would require several hundred kilowatts of power routinely and a gigawatt of power during an engagement, according to the study; space-based nuclear reactors appear to be necessary, in spite of President Reagan's assurances that the SDI would be nonnuclear. Such weapons also call for extremely accurate sensing, tracking and pointing in order to hit a target that might be thousands of miles away. Some systems would need to be retargeted 10 times per second. The free-electron laser would require mirrors in space twice the size of those in today's largest ground-based telescopes, raising special concerns about survivability in a nuclear environment. Advanced adaptive optics, in which mirror segments are quickly moved to compensate for atmospheric distortion, or optical phase-conjugation systems will also be needed.

The necessary technology is still far short of the goals. The panel emphasizes that improvement in a technology by even a single order of magnitude requires that many obstacles be overcome. Although the APS report studiously avoids addressing whether a nationwide defense will ever be possible, or whether the SDI Organization is doing a good job, it concludes that no deployment of a directed-energy weapon will be possible until the next century, at the very earliest.

The idea of having the APS study the subject was originally pushed by Robert E. Marshak of the Virginia Polytecnic Institute, a former APS president, as a way of resolving the technical issue of feasibility. The panel had access to classified materials, and its 420-page report was endorsed by a review panel chaired by George E. Pake of the Xerox Corporation. Work was completed last September, but then the study was tied up for seven months ("too long," according to Patel) in classification review, first at the Pentagon's SDI Organization and then in the office of the Secretary of Defense. "Small but significant" deletions had to be made but did not affect the principal conclusions.

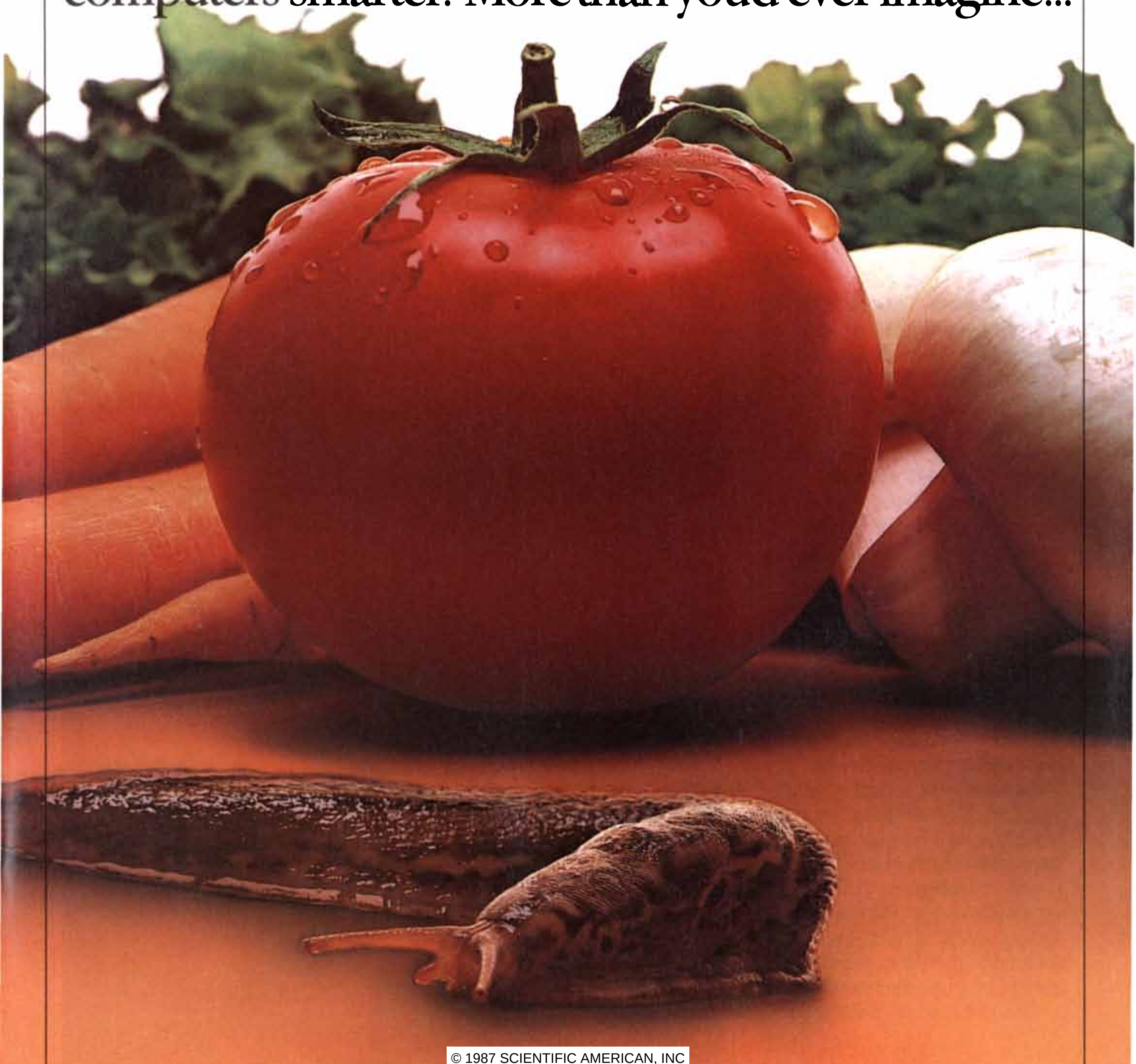
While the APS report made its seven-month journey to daylight, Secretary of Defense Caspar W. Weinberger floated the idea of deploying "phase 1" of a missile-defense system in the early 1990's. This system would not employ directed-energy weapons at the outset; instead it would rely on kinetic-kill vehicles—such as "smart rocks"—to destroy warheads by force of impact. Some would be launched from satellites. Patel warned, however, that any system employing space-based components is vulnerable to antisatellite attack or other countermeasures. Furthermore, he notes, even a kinetic-kill system must depend on directed-energy devices to distinguish warheads from decoys: substantial advances are needed even for this less demanding function. Jeremiah D. Sullivan of the University of Illinois at Urbana-Champaign, a panel member, said "one shouldn't gamble now" that directed-energy weapons will become available in the future.

The gloomy prognosis seems likely to add to the SDI's difficulties on Capitol Hill. The Administration has asked for $5.8 billion for the program in 1988, but there is little chance that Congress will provide it all. One view has it that the SDI will be lucky to get any increase above this year's level of $3.53 billion.

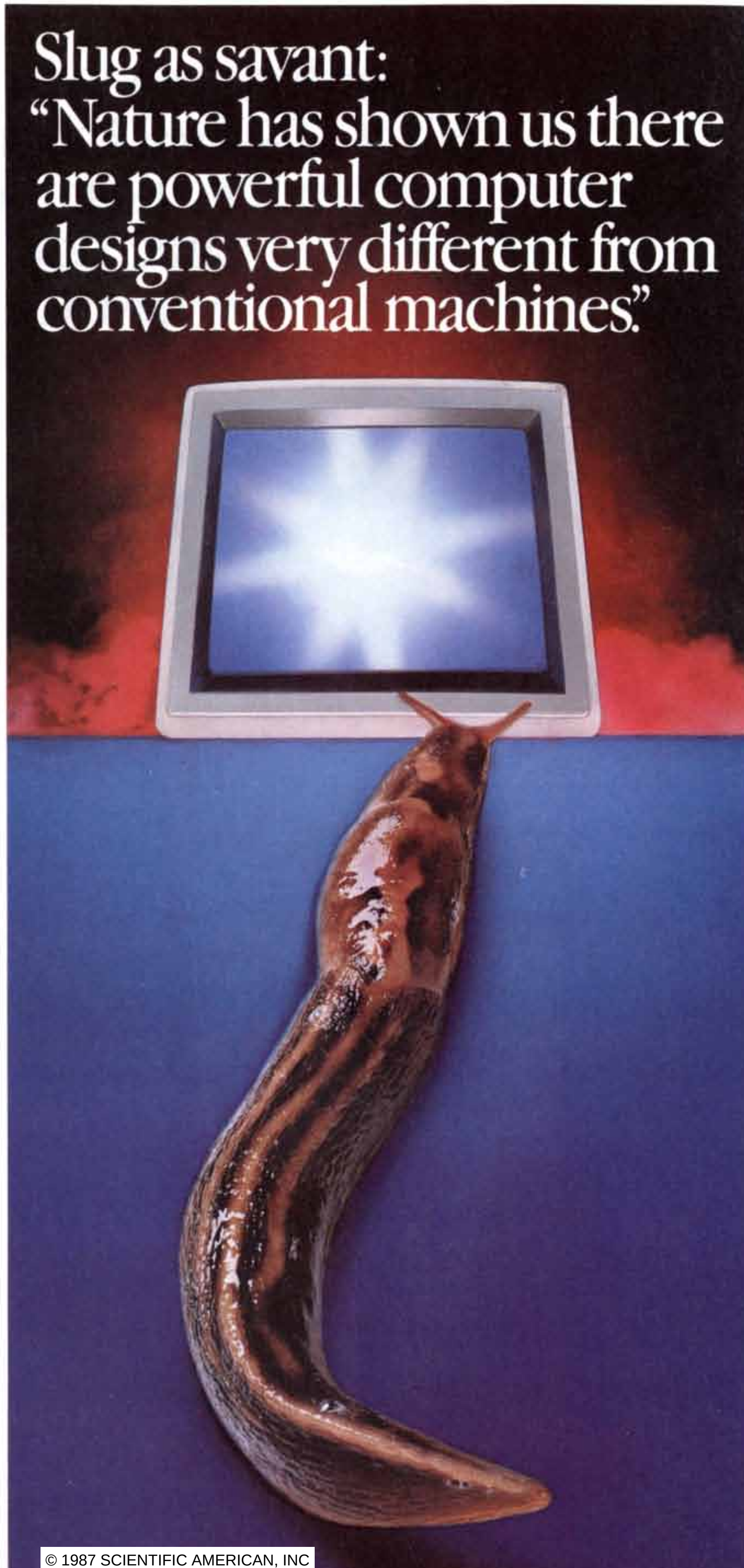
Yet the initial response of the SDIO was muted. The organization called the individual technical assessments in the APS report "objective" and the analysis "astute," but it questioned some technical assumptions and said the overall conclusions were "subjective and unduly pessimistic."

Supernova Neutrinos

Supernova 1987A, the brightest stellar explosion observed in modern times, has heated up debate over a crucial question of modern physics: What is the mass, if any, of the most abundant particle in the universe, the neutrino? Cosmologists have long speculated that if neutrinos are sufficiently massive, they might brake and even reverse the present expansion of the universe. The same elusiveness that enables neutrinos to pass through stars



AT&T Bell Laboratories: Technology for the real world



The common garden slug loves the enticing odors of carrot, tomato and mushroom. But it hates garlic. When scientists at AT&T Bell Laboratories "spike" these favorite foods with garlic, what happens? The slug learns. It alters its memory of the foods it once loved and avoids or rejects them.

Insights gained from studying simple central nervous systems like the slug's point to a dramatically new approach to computing. An approach that promises to make computers faster, smarter and easier for people to use.

Why study slugs? Though the slug is no Einstein, its brain's limited ability to learn—to associate new information with existing memories—makes today's most powerful computers seem primitive.

And the slug, with its neural networks comprised of a mere 500,000 nerve cells or neurons, is much less complicated to work with than people or other animals.

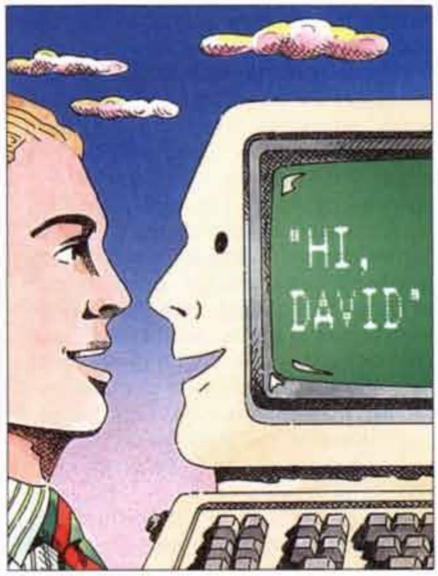
Microchips that mimic the brain

On functioning computer chips, microelectronics researchers have built prototype electronic neural networks.

Like biological networks of brain nerve cells, these electronic circuits use associative memory to relate incoming information to memories already stored. So they can cope with information filled with errors or ambiguity. And they can deal with "messy" information, collecting scattered facts to recognize and remember from incomplete details, much as the brain does.

One test chip, containing 54

In the future, people and their computers will have a much friendlier working relationship.



"neurons" in such a "neural network," can recall memories from imperfect data within a few millionths of a second—selecting "James Lynn" from among several stored names as the correct response to the input "Jim."

Getting up to speed

By studying simple central nervous systems like the slug's, scientists at Bell Labs are also gaining valuable insights into another brain function, parallel processing. It offers an answer to a physical limitation of today's computers—speed.

Step-by-step computing can only process information one piece at a

Today's computers can only process neatly-stored information. A little human ability to deal with messiness might actually make them a whole lot smarter.

time. Parallel processing, the ability to perform several functions simultaneously, speeds things up. And the more things done together, the faster the whole job gets done.

"Thinking" computers

Where is this research into associative memory and parallel processing leading? To "thinking" computers that make good judgments as well as good calculations. Computers that can perceive and learn in an imperfect world, much as people do.

In the future, working with computers will be more like working with people. The machines will understand and respond to human speech—even recognize the person addressing them.

Taking the longer view

Research scientists at AT&T Bell Laboratories are expected to take the longer view. To look beyond the impact of technology on the next quarter or the next year, into the next century.

It is this perspective that has produced seven Nobel Prize winners for Bell Labs. Some 21,000 patents, an average of more than one a day. And a legacy of achievement, from the transistor and the laser to lightwave communications and the digital computer.

This longer view ensures that the technology built into all AT&T products can evolve and adapt to the changing needs of the real world. Making information easier to obtain and use for everyone.

and planets unimpeded has made it very difficult to gauge their mass in the laboratory.

There are actually three kinds of neutrino—the electron, muon and tau neutrino—and according to theory a supernova can produce them in two distinct pulses. The first pulse occurs when electrons and protons in the collapsing core of the star merge to form neutrons and electron neutrinos. As the electron neutrinos escape, the core collapses still further and emits a pulse of all three types of neutrinos and their antimatter counterparts.

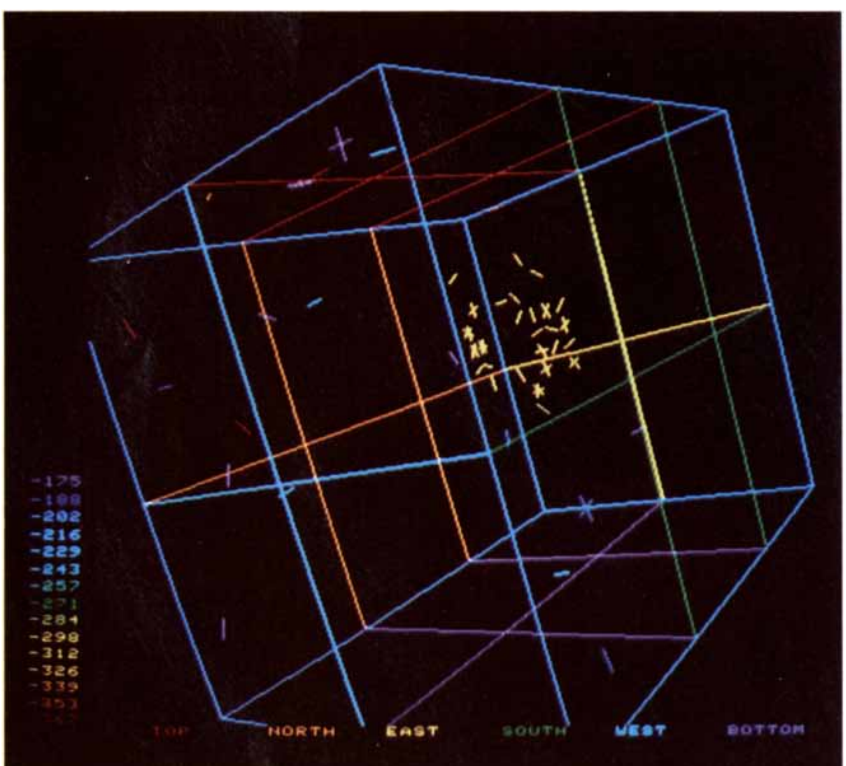
Most investigators think it was the second and more powerful pulse from 1987A that, about three hours before the supernova appeared, triggered signals in two particle observatories. The IMB observatory, buried in a salt mine near Cleveland, Ohio, and operated by the University of California at Irvine, the University of Michigan and the Brookhaven National Laboratory, recorded eight "events" over a period of six seconds. The Kamiokande II observatory, situated in a mine near Kamioka in Japan and run by Tokyo University, the National Laboratory for High-Energy Physics of Japan, the University of Niigata and the University of Pennsylvania, recorded 11 events occurring over an interval of 13 seconds.

All the events seem to have been caused by electron antineutrinos—the antimatter twins of the electron neutrinos—of various energies. Since the twins are believed to have the same mass, physicists knew that a fairly simple calculation could establish an upper limit on the rest mass of the electron neutrino. The calculation involves analyzing the degree to which higher-energy electron antineutrinos preceded lower-energy ones. If the particles had a single rest mass, those with higher energies should have traveled faster than those with lower energies; the greater the rest mass is, the greater the difference between the arrival times of the high-energy neutrinos and those of the low-energy ones should have been.

John N. Bahcall of the Institute for Advanced Study in Princeton, N.J., and Sheldon Lee Glashow of Harvard University were the first to publish an estimate. They state in *Nature* that the electron neutrino's mass must be less than 11 eV (electron volts). Bahcall and Glashow assert that their result "confirms the view that electron neutrinos do not constitute the major component of the matter density of the Universe." Recently several groups, following painstaking measurements of electron neutrinos produced in the laboratory by radioactive decay, have reported an upper limit on their mass of about 20 eV, not quite low enough to rule out the possibility that they could cause the universe to contract.

Alfred K. Mann of the University of Pennsylvania and the Kamiokande II observatory notes that other investigators analyzing the same data as Bahcall and Glashow have derived estimates ranging from less than 4 eV to more than 20 eV. These divergences, he says, reflect the fundamental uncertainty that all the estimates build on. "The crucial issue," Mann explains, "is that we know the arrival time of the neutrinos but not the departure time." The most "rigorous" handling of this uncertainty, he contends, provides an upper limit of about 20 eV.

John C. van der Velde of the University of Michigan and the IMB project suggests that if neutrino detectors gather more data from other supernovas, "then maybe we'll be able to nail this down." Van der Velde points out, however, that even if the electron neutrino proves to be not massive enough to "close" the universe, there remains unresolved the question of whether the muon or tau neutrino might accomplish the job.



ANTINEUTRINO from supernova 1987A creates flashes of light (*yellow symbols*) in the IMB particle observatory in this computer simulation. Other symbols show extraneous signals. The IMB consists of a tank of water lined with photomultiplier tubes. The number of slashes in a symbol is proportional to the signal's intensity. Colors of the symbols show arrival times, which can be read from the scale of nanosecond intervals at the left.

Global Questions

Is the ozone layer being depleted by chlorofluorocarbons? Is the greenhouse effect raising the air temperature and melting polar ice? What impact will eradication of tropical rain forests have on climate and on the viability of life? Such questions demand careful study of an imposing experimental subject: the planet earth.

Investigators from many disciplines and many parts of the world are eager to take on the task, and now plans are being made for a multimillion-dollar, multidecade International Geosphere-Biosphere Programme that could become the largest scientific project ever undertaken. William E. Gordon, foreign secretary of the U.S. National Academy of Sciences, says the IGBP is "the most important thing scientists will do this century." In April the academy appointed a committee to oversee U.S. participation.

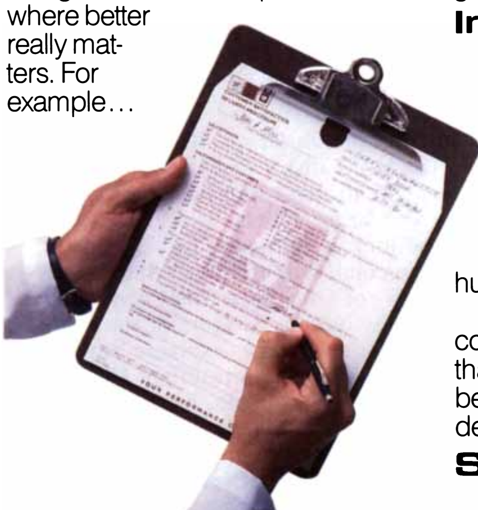
The effects of human activity on the atmosphere, hydrosphere, geosphere and biosphere "will significantly alter our habitat within a few human generations," according to a National Aeronautics and Space Administration study. The implications of that now widespread view inspired a "realization that science had to assert itself,"

We confidently invite you to give your brand new Buick its 201st inspection.

Buick Century

The first 113 quality inspections happen while your Buick is being built. But that's just a good beginning. Once it reaches your dealer, 87 additional items must be checked before it can be subjected to the most important inspection of all: yours.

And that's just one of the important things we pay attention to. Things that make us a place where better really matters. For example...



Underneath

The road of life has its share of potholes, washboards and bumps. Fortunately, it also has an occasional sweeping curve or twisting, tree-lined lane. Century is prepared to deal with it, either way.

Its independent front suspension, computer-selected springs and shock absorber valving deliver the kind of smooth, effortless ride for which Buick has long been admired, while taking very good care of business when the going gets brisk and sporting.

Inside

When you sit behind the wheel of a Century, you find that the controls are where you want them for ease of operation. That's because of our dedication to ergonomics — the science of matching machine design to human design.

Other reasons for Century's comfort include things like seats that provide real support without being hard or harsh. Aerodynamic design to cut down on wind noise.

Satisfaction

Century is the best-selling Buick in history. The reasons are many. But one of them surely is the way this automobile manages to blend today's technology with traditional grace and comfort.

It is a front-wheel-drive car, with an engine controlled by microchips. It was designed with the aid of computers and wind tunnels for superb aerodynamics. It represents what can be done with the application of the latest technology.

But all this electronic wizardry is there to fulfill one simple, single function: to provide satisfaction. A car that starts when you want it to, has the kind of power you expect and provides quiet, solid comfort on the road.

If that sounds like the automobile you've been looking for, buckle up and see your Buick dealer.

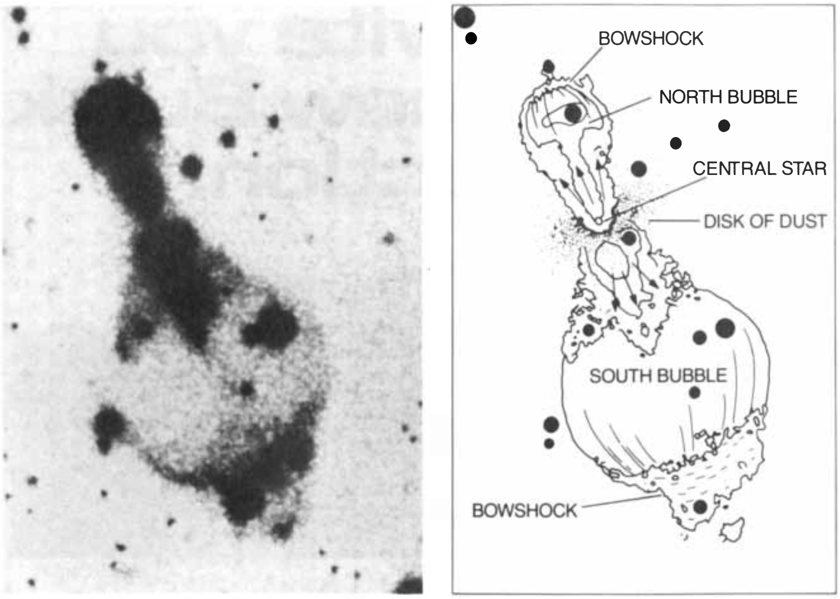
For more information, a brochure or a test drive, call 1-800-87-BUICK (1-800-872-8425).

GM 6/60 QUALITY COMMITMENT PLAN

See your dealer for the terms and conditions of this limited warranty.

Where better really matters.

B U I C K



DYING STAR ejects two bubbles of matter in a charge-coupled-device image made by Bo Reipurth at the European Southern Observatory in La Silla, Chile. The larger bubble is .8 light-year long. The interpretation at the right is based on data from four telescopes.

says James J. McCarthy of Harvard University, chairman of a committee established by the International Council of Scientific Unions to coordinate various countries' efforts.

One focus of the IGBP will be the concentration of carbon dioxide in the atmosphere. The gas will probably reach twice its preindustrial level by the middle of the 21st century, mainly because of the burning of fossil fuels and the destruction of tropical rain forests. The consequent increase in the greenhouse effect is considered likely to cause a global warming of two or three degrees Celsius; indeed, recent studies may have detected the beginning of such a trend.

Yet the phenomenon is influenced by factors far beyond human economic activity. Marine plankton sequester much of the anthropogenic carbon dioxide in deep ocean sediments, where it may remain for millions of years. Plankton populations are limited by nutrients such as nitrates, which are most plentiful where deep ocean water wells to the surface. Such upwelling depends on ocean currents, which are driven in part by winds, which in turn are affected by radiative properties of the atmosphere. Clearly the carbon dioxide cycle is complex, and it is far from completely understood. Even less is known about other important atmospheric gases such as methane.

In order to gather data from so many spheres a number of national and international projects will probably be subsumed under the IGBP. The National Science Foundation has initiated a "global geosciences" program, for which it is seeking $64 million in 1988. NASA plans several compatible geoscience missions including TOPEX, a joint U.S./French satellite to be launched in 1991 that will measure ocean currents. Other planned satellite missions will measure oceanic plankton abundance and sea-surface winds. There is informal support for a NASA "Mission to Planet Earth" that would include the accelerated deployment of geostationary platforms.

Both the White House and Capitol Hill appear to be susceptible to grand science projects, and a wide range of disciplines are crowding in. There has been jostling, including "some disagreement" about how much solid-earth geophysics should be included, according to C. Barry Raleigh, director of the Lamont-Doherty Geological Observatory. Is the IGBP, then, no more than a clever packaging of science? McCarthy acknowledges that the newly formed U.S. IGBP committee will have to convince skeptics that the IGBP is "more than just a marketing effort."

For now enthusiasm is high. NASA and the NSF are committed. The National Oceanic and Atmospheric Administration, which has long conducted climate measurements, is "seriously considering" its role, according to J. Michael Hall, director of NOAA's office of climatic and atmospheric research. The Department of Energy, whose efforts to study atmospheric carbon dioxide have been severely criticized, is reconsidering its efforts; the Environmental Protection Agency, the National Park Service and even the Navy might join in.

Physical Sciences

Not with a Bang...

A star such as the sun that lacks the mass necessary to explode into a supernova dies in a much less sudden and violent way. As hydrogen in its core fuses into heavier elements, convection currents make the outer layers of the star swell, transforming it into a red giant. Eventually the outer layers are blown away and the dim, shrunken core—a white dwarf—is exposed.

An odd, hourglass-shaped object in the constellation Puppis that is visible, like Supernova 1987A, in the Southern Hemisphere represents the climactic and most poorly understood event in this drama: the red giant's rapid loss of its outer layers. Bo Reipurth of the European Southern Observatory in La Silla, Chile, has constructed a detailed picture of the object, which is designated OH231.8 + 4.2, on the basis of readings from four telescopes. The star itself is concealed within a disk of dust at the object's center; from each side of the disk two bubbles of luminous matter are rapidly expanding.

The larger bubble, which is .8 light-year long, is dilating at 140 kilometers per second. Reipurth notes that the rapidity of the transition—compared with the billions of years of the average star's lifetime—explains why it is not witnessed more often. He estimates that the bubbles began to form some 1,500 years ago; in several thousand years they will have ceased expanding and will disperse into what is called a bipolar planetary nebula.

Dozens of fully formed bipolar nebulas, some with white dwarfs discernible at their center, have already been spotted. One theory on their formation holds that they evolve from red giants ringed by a disk: as the aging star loses its outer layers, the matter expands at right angles to the disk because that direction offers the least resistance. Reipurth says OH231.8 + 4.2 seems to match this model.

Why do disks develop around red giants in the first place? The question touches on a larger one: to what extent does OH231.8 + 4.2 foreshadow the sun's fate? It is generally agreed that disks form when gas rises from the surface of a red giant, blown by the equivalent of the solar wind, and condenses while the star is still relatively stable. Mark R. Morris of the University of California at Los Angeles asserts that the gravitational pull of a companion star is the likeliest mechanism for causing condensed gas to fall into a particular orbit and form a disk rather than a roughly spherical envelope. On this basis Morris proposes that OH231.8 + 4.2 "probably" is a binary star and not a single star like the sun.

Benjamin M. Zuckerman, also of U.C.L.A., disagrees with Morris that disks—and so too bipolar nebulas—probably arise only from binary stars. A disk might form around a single star, Zuckerman suggests, if the star rotates rapidly or has a powerful dipolar magnetic field. "We just don't know yet," he observes.

Limited Occupancy

Space scientists are close to rebellion over the National Aeronautics and Space Administration's new two-phase space-station plan. The program now provides for an initial habitation that would support much less research than had been hoped for. Occupancy would be delayed until 1996, which is two years after the date foreseen when the project was approved by President Reagan in 1984.

The new blueprint (which is not the first downgrading of the project) was forced on NASA by estimates indicating that the cost of the version of the station announced last year had risen from $8 billion to $14.5 billion in 1984 dollars—about $16 billion in 1987 dollars. The White House intervened early this year and opted for a two-phase approach, phase 1 of which would cost $12.2 billion in 1984 dollars. Phase 1 has only one laboratory for U.S. investigators and one living module; it omits a previously planned co-orbiting experiment platform and satellite-servicing bay. It also has only two attachment points for scientific instruments. The power supply would be limited to 50 kilowatts, 30 kilowatts of which would be needed for routine housekeeping functions.

Even before the latest modification was announced there was serious dissent among space scientists over NASA's plans. The backlog of shuttle launches caused by the *Challenger* disaster will delay dozens of scientific missions, and a task force that advises NASA on scientific uses of the space station had noted "with deep regret" that the space-station schedule was "beyond the operative horizons of the present generation of space scientists." Peter M. Banks of Stanford University, former chairman of the task force, even proposed that NASA use a contemplated new Heavy Lift Launch Vehicle to deploy a single large module so that the program could start before 1996.

Banks's idea was immediately rejected by NASA, and he resigned soon thereafter. He says the new two-phase space-station plan confirms his view that NASA is "not honoring its commitment" to man-assisted space science; he would prefer to see early deployment of small, tended platforms.

Others think the space station is a bad idea altogether. One is Thomas M. Donahue, chairman of the National Research Council's Space Science Board. Donahue wrote to congressional committees in April urging that NASA be forced, if necessary, to purchase expendable rockets so that unmanned space-science missions could be flown. Donahue believes there is no urgent scientific justification for a space station and thinks its budget is eating into space science. He adds: "We risk repeating the experience of the *Challenger*."

Reaction to NASA's plan in Congress was mixed, and the reduced power provoked skepticism. One option being discussed would exclude foreign users of the station until power had reached the previously planned 87.5-kilowatt level. Detailed congressional scrutiny will be deferred until a report on cost being drawn up by the National Research Council is completed in September. Negotiations with foreign space-station partners should, however, become somewhat easier: U.S. agencies recently agreed on a new proposed text for the formal intergovernmental agreement that goes some way toward meeting European objections to unfettered U.S. military access.

The Impossible Dream

More than 20 years ago Edwin T. Jaynes of Washington University in St. Louis and Frederick W. Cummings of the University of California at Riverside considered how a single excited atom would behave in a resonant metal cavity "tuned" to a particular electromagnetic frequency. Such an atom can emit a photon, or quantum of electromagnetic energy, that establishes a radiation field, from which the atom can then reabsorb a photon. The two theorists predicted that the frequency of such photon exchanges would be influenced by the presence of other photons in the cavity and so would affect the overall likelihood that the atom would be found in an excited state.

Jaynes and Cummings showed that as more photons are included in the cavity, the number of photons interacting with the atom becomes indeterminate. This indeterminacy implies there is no longer a definite photon-exchange frequency but rather a distribution of such frequencies, and the probability of finding the atom in an excited state drops. If the field and the atom interact long enough, however, the field can eventually manifest itself to the atom as a discrete number of photons, and the probability of finding the atom in an excited state increases correspondingly.

At the time they outlined their theory Jaynes and Cummings considered their postulated system to be impossible to realize in practice. But times change, and now Gerhard Rempe and Herbert Walther of the University of Munich and Norbert Klein of the University of Wuppertal have actually re-created the theorists' thought experiment: they stimulated an excited rubidium atom to emit a photon of microwave radiation in a supercooled, superconducting resonant cavity.

To make the atoms enter the cavity one at a time the German workers passed a collimated beam of rubidium atoms through a series of rotating slitted disks, adjusting the rotational speed so that only atoms traveling in single file and at a certain velocity could run the gauntlet into the tiny entrance hole of the cavity. Just before entering the cavity each atom was excited by a laser to a high-energy state from which it could emit a microwave photon. A detector placed outside the cavity's tiny exit determined whether an atom leaving the cavity had left a photon behind.

So little energy leaks out of the resonant cavity that a photon emitted by an excited atom can loiter in the cavity long enough (about two milliseconds) to be absorbed either by the same atom or by another one following it. Hence an excited atom enters the cavity, where it can emit and absorb photons repeatedly. If the emerging atom is not in an excited state, the cavity field for the next atom has been enriched by one photon.

Since the investigators could regulate both the average number of photons in the cavity (by adjusting the rate at which the atoms entered the cavity) and the atoms' length of stay in the cavity (by adjusting the velocity of the atoms), their experimental setup was ideal for checking the validity of the Jaynes-Cummings theory. As the experimentalists report in *Physical Review Letters,* the results are in agreement with the predicted behavior.

Central Heating

By compressing minute samples of iron between two diamonds and striking thin iron targets with high-speed projectiles, investigators have arrived at unprecedentedly high estimates of the temperatures prevailing in the earth's core. Writing in *Science,* groups working at the University of California at Berkeley and the California Institute of Technology tell how they derived the temperatures from measurements of the melting point of iron—the main constituent of the core—at immense pressures.

In their reasoning the workers relied on several well-established features of the earth's core. Seismic studies have shown that the outer core, which meets the overlying mantle at a depth of about 2,900 kilometers, is molten. Hence the melting point of iron under the pressures calculated to prevail at the top of the outer core sets a lower limit for temperature there. The inner core, which extends from a depth of about 5,200 kilometers to the center of the earth, at 6,370 kilometers, is solid. The temperature at the interface of the inner and the outer core must equal the melting point (which increases greatly with pressure and therefore with depth).

Direct measurements of high-pressure melting done earlier had been limited to about 20 billion pascals, or about 200,000 times atmospheric pressure—equivalent to conditions a mere 600 kilometers down. A combination of technologies enabled the workers, led by Raymond Jeanloz of Berkeley and Thomas J. Ahrens of Caltech, to extend the measurements to 250 billion pascals, thereby simulating conditions well within the core.

To reach pressures of up to 100 billion pascals the Berkeley investigators compressed iron in a diamond-anvil cell, a device in which a simple arrangement of screws and pistons transmits pressure to a sample trapped between the faces of two diamonds. The transparency of the diamonds made it possible to heat the iron with an infrared laser and determine the sample's melting temperature at various pressures from the spectrum of emitted light. The Caltech group attained even higher pressures by employing compressed hydrogen to fire a projectile at a film or foil of iron mounted on a crystalline window. The impact simultaneously compressed and heated the sample, whose temperature was measured from thermal radiation detected through the window.

The measured relation of melting temperature to pressure showed that the temperature at the top of the outer core, where the pressure is 136 billion pascals, must be at least 3,800 degrees Kelvin. (The melting temperature of pure iron at that pressure is actually 1,000 degrees higher, but the authors reduced their estimate to take into account the effect of the impurities thought to be mixed with iron in the core.) Extrapolating the results indicated that at the top of the inner core, where the pressure is 330 billion pascals, the temperature is about 6,600 degrees. The workers offer 6,900 degrees K.—hotter than the surface of the sun—as an upper limit for the temperature at the center of the earth.

Those figures are between 2,000 and 3,000 degrees higher than past estimates based on less direct evidence. Because temperature differences drive convective motions in the earth's interior—processes that manifest themselves in the shifting of continents and the eruption of volcanoes—the geophysical consequences of an extremely hot core could be profound. The results imply that "a very high fraction of the heat in the earth's mantle originates in the core," according to Ahrens. "The core may play a larger role in mantle and crustal motions than was previously believed."

BIOLOGICAL SCIENCES

Making Antisense

DNA dogma holds that just one strand of a double-strand DNA segment codes for a genetic product. That strand makes sense to the cell's interpretive machinery, and so it is transcribed into the liaison molecule RNA; the other, "antisense" sequence merely preserves the integrity of the coding sequence. Within the past year a handful of experiments have challenged that model, and now a team of investigators at the Oregon Health Sciences University has found conclusive evidence that two genes can occupy one stretch of DNA, running side by side on opposite strands.

The investigators, headed by John P. Adelman of the Institute for Advanced Biomedical Research, were looking in rat brain tissue for copies of the gene for gonadotropin-releasing hormone (GnRH). As they explain in *Nature,* they used probes derived from transcribed strands of GnRH DNA; the probes were intended to recognize the sense strand of the GnRH gene.

Instead the probes stuck to both strands, implying that the gene sequence opposite the GnRH sequence was also being transcribed. Adelman and his colleagues then examined a spectrum of rat tissues for evidence of this antisense transcription. Because they observed most of it in the heart, they called the antisense gene SH, for "sweetheart."

Antisense transcription had been described in bacteria, but until last year it was unknown in higher forms of life. Then experiments with insect cells and with mammalian cells in tissue culture showed that genes on opposite strands may overlap. These reports, however, were generally considered exceptional cases in unusual cell specimens. In contrast, Adelman's research was done with ordinary laboratory-rat tissue, and the GnRH and SH genes coincide on the DNA instead of overlapping only partially.

Adelman says no one knows yet how prevalent cohabitating genes are, but he suspects that other researchers' experimental "artifacts" may turn out to be genes that had gone undetected. Many experiments mapping gene locations on chromosomes may have to be repeated to ascertain that a genetic partner has not been overlooked.

This new twist in the double-helix story may also have implications for evolutionary strategies. GnRH is a hormone necessary for sexual development. Its gene is consequently very stable, because mutation or deletion would render an individual sterile and prevent the gene from being passed on. Any other gene in the GnRH gene's territory will share in this stability whether or not it is as important as the one for the reproductive hormone. Hence the SH gene, whose function is still a mystery, has ridden "piggyback" on the evolutionary success of the GnRH gene. If other antisense DNA strands do in fact make sense, they too may share in and also help to shape the evolutionary fate of their next-strand neighbors.

Old Memories

Is memory loss a universal corollary of aging? For old people who are in vigorous health some kinds of memory loss are by no means inevitable, results presented at a recent meeting of the American Academy of Neurology suggest.

Elisabeth Koss and her colleagues at the Laboratory of Neurosciences of the National Institute on Aging tested the ability of 60 men, ranging in age from 20 to 85, to retain specific kinds of information. All the subjects were in excellent health; in particular the investigators excluded people suffering from conditions that can lead to impaired brain function, such as high blood pressure. (Among candidates for the study who were more than 60 years old the rate of rejection was between 60 and 80 percent.) The subjects were also well educated and active. Hence the results define not average capacities but upper limits: the degree to which memory can be preserved, given the right physical constitution and environment.

Earlier workers had found that age affects two kinds of mental abilities

differently. Verbal ability, that is, the capacity to understand and employ language, does not seem to decline in healthy older people. Visuospatial ability—the capacity to manipulate and represent objects in space—does falter with age. Koss and her colleagues found that the same pattern of a spared verbal faculty and an impaired visuospatial one extends to memory. Age, that is, had no bearing on performance in tests of verbal memory, in which the subjects were asked, for example, to remember a series of words or the gist of a story. Only when the workers tested visuospatial memory, for instance by showing a drawing to a subject and then asking him to reproduce it from memory, did performance fall with age.

Because memory loss in healthy older people seems to affect only specific kinds of information, the workers speculate it may be a by-product of a different kind of decline. Instead of revealing a failure of memory mechanisms, trouble remembering a picture might stem from an inability to assimilate it in the first place, because of slowed mental processing or a reduced ability to concentrate—both well-established features of aging. Attending to, remembering and reconstructing a pattern or arrangement may tax those abilities more than verbal tasks do; the visuospatial task may also require a higher level of motivation, which also declines with age. The more extensive memory loss seen in some old people, Koss and her colleagues think, generally results from disease rather than from normal aging.

Frog Went a-Courtin'

Among gray tree frogs, inhabitants of woodlands throughout the eastern U.S., females do the choosing. Sitting by a pond on a dark summer night, a female listens to a chorus of males uttering urgent, trilling calls. She selects a mate in a straightforward manner: she hops up to him. To make sure he gets the message, she may nudge him with her nose.

On what does she base her choice? Does she automatically hop to the loudest (usually also the nearest) male, as some biologists have believed, or is she more discriminating? Georg M. Klump and H. Carl Gerhardt of the University of Missouri report in *Nature* that females of at least one species of gray tree frog, *Hyla versicolor,* listen to males' calls with a distinct set of aural priorities in mind.

The investigators tested the relative preference of the female *H. versicolor* for three properties of male calling: the duration of a single call, the rate of the trills or pulses that make up a single call and the rate at which calls are sounded. One female at a time was put in a chamber with two speakers, each of which emitted a different artificially produced male call. Subjects indicated their choice by hopping over to or onto a speaker.

Pulse rate was found to be by far the most important criterion for a female *H. versicolor*. Earlier observations had indicated that pulse rate is the only feature by which the frog (or a human observer, for that matter) can distinguish between a male *H. versicolor* and a male of another species of gray tree frog, *H. chrysoscelis*. A female *H. versicolor* will approach the call of a male *H. chrysoscelis,* according to Klump and Gerhardt, only if she can hear no slower-trilling *H. versicolor* calls, however faint and distant. She thereby avoids what Gerhardt calls "the biggest mistake she can make" from an evolutionary point of view: hybrid offspring of the two species are sterile.

Duration of call is the next most important factor. Females favor longer calls over shorter ones even if the latter are emitted more frequently. Klump and Gerhardt cite research done at the University of Connecticut that suggests male gray tree frogs expend energy more rapidly when they emit fewer but longer calls. The Connecticut workers also found that as the number of males competing for females in a single area increases, the duration of their calls also increases. Indeed, the choosiness of the females, Gerhardt says, seems to "push the males to their physical limits." He and Klump conclude that a male's ability to rise to the challenge may reflect his "energetic investment...in courtship" and, indirectly, his physical fitness.

Some males, Gerhardt notes, win a mate without a single call. The so-called silent male merely positions himself near a noisy competitor and, when a female approaches, quickly mounts her.

FEMALE TREE FROG shows her preference for a particular male mating call by climbing the speaker that emits it. The experiment was done at the University of Missouri.

MEDICINE

Doctored Diagnosis?

Not-for-profit hospitals are supposed to offer low-cost care for all, particularly the medically indigent. Do they? Or do they soak up billions of dollars' worth of tax breaks merely to give physicians a subsidized place to practice medicine? Officials of not-for-profit hospitals are smarting over a recent study asserting that the second account is correct. The debate gains in intensity as Federal and state legislatures shape budgets and look at tax law. "Now we're having to defend our existence," says one not-for-profit hospital lobbyist.

"Who Profits from Nonprofits?" by Regina E. Herzlinger and William S. Krasker, was published in the January–February issue of *Harvard Business Review.* Herzlinger is a professor of business administration at the Harvard Business School; Krasker, formerly at the school, is now with Salomon Brothers. The study analyzes the financial performance of 14 major for-profit, investor-owned hospital chains and not-for-profit hospital chains in 1977 and 1981. Herzlinger and Krasker attempted to compare profit-making and not-for-profit institutions on equal terms by adjusting for tax differences. The workers conclude that the not-for-profit chains "have not fulfilled their social promise." Although not-for-profit chains receive social subsidies, mainly as tax exemptions, according to Herzlinger and Krasker they are neither more accessible to the uninsured nor less expensive. The discipline of the market, they maintain, ensures that for-profit hospitals "produce better results for society and require virtually no social investment."

Critics contend that the study has profound methodological flaws—primarily in sampling—that lead it into error. An exponent of this view is Bradford H. Gray, director of a study published last June by the Institute of Medicine. The IOM agreed with earlier work in finding that for-profit hospitals charge more than not-for-profit ones and also provide less free care to those unable to pay.

Uwe E. Reinhardt of Princeton University, reputedly a champion of for-profit hospitals, judges the Herzlinger study to be "empirical research bent toward the authors' predilections." Lawrence S. Lewin, a prominent investigator in the field, says he is "saddened" that Harvard lent its name (when the business school publicized the study at a press conference) to "sloppy work." Arnold S. Relman, editor of the *New England Journal of Medicine,* has written that the not-for-profit-hospital sample in the study "neglected the hundreds of larger, freestanding nonprofit hospitals where most teaching, clinical research, and free care take place."

Herzlinger says her critics misunderstand her "unique" methods of accounting. Observing that more not-for-profit institutions are becoming organized into chains, she says critics' complaints that her sample under-represented stand-alone not-for-profit hospitals confirm their "business naiveté." She denies any financial interest in the for-profit sector and criticizes earlier major studies that were sponsored by a consortium of not-for-profit hospitals: "They have to keep their clients happy." She also dismisses the IOM study as "flawed," saying it was mostly not the work of academic experts. The argument over the for-profit and not-for-profit institutions seems likely to continue.

Cancer: Illusory Progress?

Three out of 10 U.S. residents will, at present rates, be diagnosed as having cancer. This depressing statistic creates a demand for good news, a demand that the National Cancer Institute, which must sell its program to Congress every year, and the charities that support research and care are eager to meet. As a rule the good news comes as an announcement that the five-year survival rate for one or another form of cancer has increased.

A General Accounting Office study has unleashed a torrent of critical rain on this parade of victory announcements. The agency convened expert groups to discuss 12 of the commonest types of the disease, all of which, according to NCI figures, showed substantial improvements in five-year survival rate between 1950 and 1982. The GAO's panels concluded that many of the "improvements" over that period are largely the result of changes in the way survival rates are measured. Earlier and more accurate diagnosis and the spread of mass screening programs extend apparent survival rates even when there is no improvement in therapy. The disagreement is over how large these effects are. The NCI's parent, the Department of Health and Human Services, maintains they are not of much more than "academic interest" but has agreed that the NCI will routinely point them out in the future.

The GAO study acknowledges that improved treatment has led to real, if slight, improvements in survival for most of the cancers studied. There have even been "dramatic" improvements for some, such as acute leukemias and non-Hodgkin's lymphoma, that affect relatively few patients. The major killers, though, have stubbornly resisted therapeutic effort. Five-year survival in breast cancer, for example, improved from 60 to 74.6 percent over the period in question, but the GAO concludes that "a large percentage if not most" of the increase stems from measurement changes. There has been no real improvement for stomach cancer. Lung cancer, by far the commonest type, is still quickly fatal in almost all cases: chemotherapy sometimes works, but only for one not very common type.

David Korn of the Stanford University School of Medicine, chairman of the National Cancer Advisory Board, disputes the allegation that the NCI has exaggerated the advances. Korn argues that earlier diagnosis is itself an advance that will lead to increased survival. Vincent T. DeVita, Jr., the NCI's director, insists that the GAO study "does not reflect the true extent of progress" and that its opinion-based analysis limits the study's accuracy and usefulness. Edward Sondik, chief NCI statistician, says the GAO's "subjective" approach is "fraught with difficulties" and argues that "there is little if any concrete evidence on the size of these statistical effects on the data."

The NCI is committed to halving the 1980 level of U.S. cancer mortality by the year 2000. Sondik says the goal "is still a prospect" if the best treatments come to be applied widely and if recent downward trends in smoking—the leading cause of cancer deaths—continue. And that, perhaps, is the point. No one wants the NCI to close up shop but the fact remains, some preeminent clinicians note, that even billions of research and health dollars may not be able to buy the necessary changes in people's behavior.

Therapy by Transplant

Mexican neurosurgeons have reported what appears to be the first successful therapy of Parkinson's disease by tissue transplantation into the brain. The early results, in young patients, have been dramatic, and the technique is already being tested in U.S. medical centers. Much more experimental surgery will be necessary, however, to demonstrate that the procedure is safe and to find out whether it has lasting effects and is effective in older patients.

Parkinson's disease results from the degeneration of fibers that deliver a neurotransmitter, dopamine, from the substantia nigra to the neostriatum, a seat of motor control. Eight years ago animal experiments showed that the implantation of fetal substantia nigra tissue in the neostriatum could restore function, but the transplantation of human fetal tissue into patients would pose grave ethical problems. An alternative was developed in animal studies: implantation in the neostriatum of tissue from the medulla (the inner region) of the adrenal gland. Two years ago a Swedish group first treated older Parkinson's patients with adrenal autotransplants, but no significant improvement was observed.

The Mexican workers, led by Ignacio Madrazo of "La Raza" Medical Center and René Drucker-Colín of the National Autonomous University, made a slight but perhaps important change in the procedure: they graft the adrenal tissue to a site on the surface of the neostriatum within a ventricular cavity, where it is bathed by nutrient-bearing cerebrospinal fluid. They reported their first two cases in the *New England Journal of Medicine.* The seriously incapacitated male patients, 35 and 39 years old, showed remarkable improvement. The first patient, for example, had been in a wheelchair before surgery; 10 months later he was farming and playing soccer.

By now the Mexican group has done 10 more transplants in incapacitated patients of both sexes between the ages of 35 and 55. According to Drucker-Colín, all these patients showed improvement comparable to that of the first two cases. (Two of the 10 have since died, but the cause of death was not related to the surgery, he says.)

How does the tissue graft work? The chromaffin cells of the adrenal medulla ordinarily secrete epinephrine, not dopamine. Moreover, it seems unlikely that the transplanted cells extend fibers that make synaptic contact with neurons in the neostriatum. Dopamine is an intermediate, however, in the biochemical pathway that synthesizes epinephrine. There are indications that the transplanted chromaffin cells may "leak" some dopamine into the cerebrospinal fluid; the dopamine could diffuse in the fluid, either building up the store available for neurotransmission or serving to enhance the response of neostriatal cells to whatever dopamine is available from the depleted substantia nigra. In an editorial in the *New England Journal* Robert Y. Moore of the School of Medicine of the State University of New York at Stony Brook suggested another possibility: the transplanted cells secrete a factor promoting the growth of nigrostriatal nerve fibers.

Whatever the mechanism may be, Moore called the Mexican report "an important event in the history of the treatment of Parkinson's disease" and urged that the National Institutes of Health initiate a multicenter clinical trial. In April a group led by George S. Allen of the Vanderbilt University Medical Center did the first U.S. transplants, in two moderately afflicted patients in their forties. Abraham N. Lieberman of the New York University Medical Center said his group would soon begin testing the procedure, with Mexican surgeons in attendance, in severely incapacitated patients young enough to withstand the brain surgery.

Hard Knocks

Adding to the many other disadvantages facing a child who grows up poor in the inner city is an unusually high risk of a disabling head injury. So say George E. Locke and his colleagues at the Charles R. Drew Postgraduate Medical School and at the King/Drew Medical Center in Los Angeles, which draws most of its patients from poverty-stricken Watts.

At a recent meeting of the American Academy of Neurology, Locke presented the results of a study of 638 children up to 18 years of age who were treated for head injuries at King/Drew in 1983. The number (which represents only about 75 percent of the pediatric head injuries treated in that year at King/Drew) tells part of the story: it is disproportionately high, given the population served by the hospital. The workers undertook to find out why, and what the long-term effects of the high injury rate might be.

After a child with a head injury was identified in the emergency room, clinics or wards and registered with the project, a parent or guardian was interviewed about the circumstances of the injury. The case was then listed as "accidental" or "nonaccidental," a category that encompassed known or suspected child abuse, injuries inflicted by other children or adults outside the family and injuries traceable to neglect or careless supervision. Nearly 30 percent of the pediatric head injuries qualified as nonaccidental, and they were generally more severe than the accidental injuries.

The data offer insights into how an environment of violence and neglect fosters head injury in children. The number of injuries peaked in late summer, in part because of a summertime surge in the number of children hit by a motor vehicle while playing in the street—often well after dark. The largest number of injuries (more than 40 percent) were sustained in falls, often traceable to poor supervision at home or in the playground. A smaller but still disturbing fraction (about 10 percent) resulted from fights and assault.

Locke believes studies done in other areas that have experienced the social breakdown associated with poverty, high unemployment and immigration would turn up similar patterns. The evidence, he says, points to "a major public-health problem," which gains urgency from the fact that even mild head injuries can affect mental function. A study at the University of Virginia has shown that even a head injury that, for example, causes no more than a brief loss of consciousness can lead to long-lasting intellectual impairment. Locke and his colleagues found similar effects in a sample of 12 children tested weeks after mild head injury. The children did poorly in tests of visual perception, short-term memory and other capacities. School records indicated a drop in academic performance by several children. Because of the high rate of head injury in the inner city, Locke argues, already disadvantaged children "are being pushed further into the underclass."

TECHNOLOGY

Space Phone

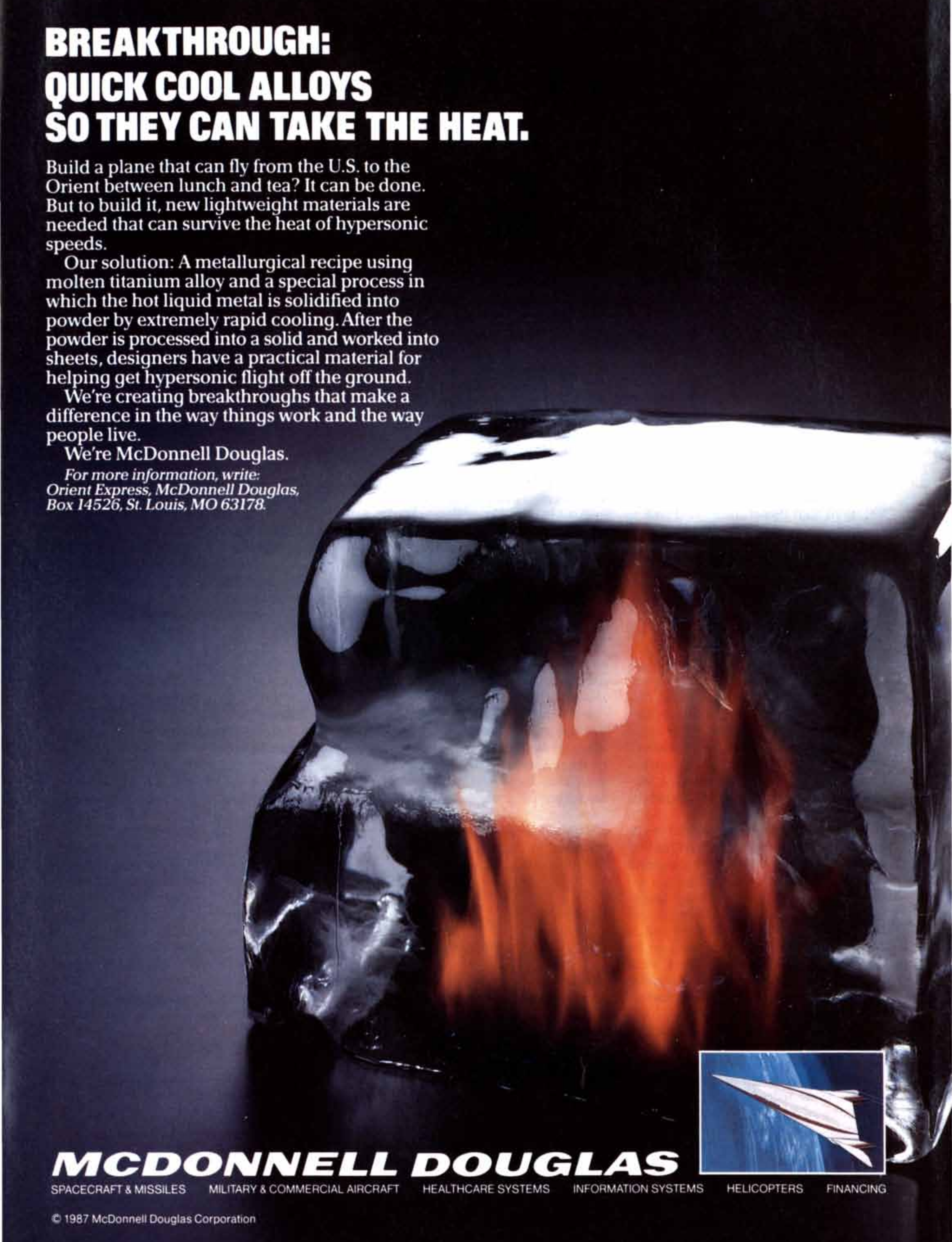
Mobile cellular radiophones, commonplace in urban areas, are useless in remote parts of the country where ultrahigh-frequency antennas are few and far between.

A high-tech solution is in the offing: satellite relay. Development of an initial satellite system for mobile receivers is thought likely to cost about $400 million—not a large sum by communications-industry standards. Indeed, the market is so attractive that the Federal Communications Commission has ruled that a consortium will be selected to develop a system, operating in the so-called *L* band. Twelve companies have now put down deposits of $5 million to indicate their seriousness; a formal proposal for licensing will be made by July 27. A demonstration system exploiting spare satellite capacity could be in place within two or three years, but eventually one or more dedicated satellites will be needed.

A similar system might also be used for air-traffic control, which at present is entirely ground-based; it would enable ground controllers to talk directly to planes over the mid-Pacific, for example. Japan is interested in the idea and plans to conduct a trial of satellite-aircraft communications this summer; Canada has already approved a system for land users.

A mobile satellite-communications system faces major technical and cost obstacles. The industry judges, probably correctly, that consumers will not want large dish antennas mounted on their cars, but small antennas are less sensitive and require the satellite end of the link to be very powerful—and expensive. Furthermore, keeping an antenna pointed accurately at a small spot in the sky while the vehicle makes turns and climbs hills is no small feat.

The Jet Propulsion Laboratory of the National Aeronautics and Space Administration, which is working on advanced technology that might later be taken up by industry, has begun field tests of two types of steerable antenna. One, which might be suitable for trucks, is a mechanically controlled flat array that could fit into something the size of a cake tin. The initial results have been encouraging, according to Elisabeth J. Dutzi of NASA's communications division. For


BREAKTHROUGH:
QUICK COOL ALLOYS
SO THEY CAN TAKE THE HEAT.
Build a plane that can fly from the U.S. to the Orient between lunch and tea? It can be done. But to build it, new lightweight materials are needed that can survive the heat of hypersonic speeds.
Our solution: A metallurgical recipe using molten titanium alloy and a special process in which the hot liquid metal is solidified into powder by extremely rapid cooling. After the powder is processed into a solid and worked into sheets, designers have a practical material for helping get hypersonic flight off the ground.
We're creating breakthroughs that make a difference in the way things work and the way people live.
We're McDonnell Douglas.
For more information, write: Orient Express, McDonnell Douglas, Box 14526, St. Louis, MO 63178.
MCDONNELL DOUGLAS
SPACECRAFT & MISSILES MILITARY & COMMERCIAL AIRCRAFT HEALTHCARE SYSTEMS INFORMATION SYSTEMS HELICOPTERS FINANCING
© 1987 McDonnell Douglas Corporation

automobiles, NASA is investigating a more advanced and expensive technology, phased-array antennas, which can "steer" a beam electronically (see "Phased-Array Radars," by Eli Brookner; SCIENTIFIC AMERICAN, February, 1985). These could be fabricated as disks two or three inches deep that would fit unobtrusively within the roof of a vehicle. Simple omnidirectional antennas have been investigated by Hughes Communications Services, Inc., one of the major consortium members. The devices are cheap and do not have to be steered, but they are more susceptible to interference caused by echoes.

JPL engineers are also developing circuits that exploit redundancy in the harmonic content of the voice to compress speech signals into a mere 4,800 bits of data per second and thereby reduce pressure on bandwidth, a major limiting factor. The signal is electronically reconstituted by the receiver into something that with some luck will not sound too synthetic. Another major effort is under way to develop satellite antennas that direct individual beams to discrete areas so that the same frequency can be used in widely separated regions without interference.

In the Light Spot

Last year investigators at the AT&T Bell Laboratories reported the first successful operation of an optical atomic trap: a device that relies on the forces generated at the focal point of a laser beam to trap atoms for a few seconds in a volume of about 1,000 cubic micrometers (millionths of a meter). Not content to rest on their laurels, two workers at the laboratories, Arthur Ashkin and Joseph M. Dziedzic, went after bigger game.

They have applied a similar trapping principle to capture individual viruses, bacteria and even some small multicellular microorganisms in a laser beam—apparently without harming the specimens. Moreover, once they are trapped in the beam the bacteria and larger microorganisms can be readily moved about, as if they were being manipulated by tweezers, while being viewed through a high-resolution optical microscope.

The trapping device constructed by Ashkin and Dziedzic relies on a subtle force, known as a gradient force, that arises in a laser beam of uneven intensity. If the intensity of light is greater at the center of the beam than at the outer fringes, then a particle with a high refractive index (a particle that strongly deflects light rays passing through it) tends to move toward the center of the beam. The reason is that each of the beam's light rays gives rise, as it is refracted by the particle, to a force in the direction opposite to the direction in which it is bent. Light rays in the center of the beam are refracted outward and their resultant force is directed inward, toward the center of the beam; light rays on the fringes of the beam are refracted inward and tend to force the particle out of the laser beam. Because the rays in the center are stronger than those on the fringes, the central rays exert the dominating influence and the net force is inward.

A radially nonuniform beam draws particles to its central axis only; it does not prevent them from being swept along the axis by photons traveling in the direction of the beam. To nullify such axial radiation pressure, the intensity of the beam must vary along its axis. This is readily achieved by focusing the beam into a small spot. If the beam is focused tightly enough, particles are drawn to and held where the light is most intense—at the focal point. To move the trapped particles it is necessary only to move the beam.

Ashkin and Dziedzic report in *Science* that the first biological particles to be captured were viruses. Although the investigators successfully confined viruses suspended in water within volumes measured in cubic micrometers, they were puzzled by the fact that in water samples that had been kept for several days "some strange new particles" would often get caught in the optical trap. The particles, about 100 times larger than an individual virus, seemed to be self-propelled; they also proliferated. When the workers looked through an optical microscope at one of these curious particles, they discovered that they had caught a rodlike bacterium. Encouraged by their success at trapping motile bacteria, Ashkin and Dziedzic have since gone after microscopic worms and have captured them in the light trap.

SURVEY

The New Superconductivity

Ceramic oxides superconduct at above 90 degrees Kelvin; theorists struggle to understand why as a technological tsunami begins to build.

What theory best explains why some of the materials known as ceramic oxides superconduct at temperatures that only one year ago would have been considered unbelievably high? Asked this question, a group of experimentalists at the AT&T Bell Laboratories, all of whom are deeply involved in developing the extraordinary new materials, exchanged grins. "The latest theory I've heard," Robert J. Cava said, "is that they're pieces of the supernova that have fallen to the earth." In a more serious tone another physicist observed, as his colleagues nodded in agreement: "I haven't seen a theory yet that can help me to find better superconductors."

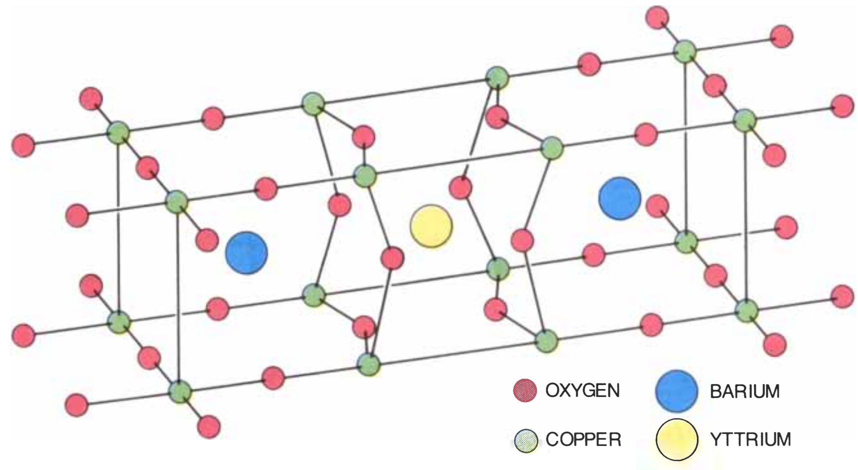
While theorists continue to puzzle over why the new superconductors work so well, bench physicists, relying on dogged trial-and-error experimentation, keep making them work better. It was only last fall that investigators at the IBM Zurich Research Laboratory initiated what all now agree is a revolution in solid-state physics by reporting that a ceramic compound of lanthanum, barium and copper oxide superconducts at temperatures as high as 28 degrees Kelvin. The previous record, 23 degrees, had stood for more than 10 years. By February several groups found that a compound of yttrium, barium and copper oxide has a T_c, or highest superconducting temperature, above 90 degrees K.

Since then other investigators have confirmed the finding and, by substituting various elements for yttrium, have discovered about a dozen more compounds that superconduct in the same temperature range. Unlike ordinary superconductors, which must be cooled with liquid helium (boiling point, 4.2 degrees K.), the new compounds can be cooled with liquid nitrogen (boiling point, 77 degrees K.), which is almost as cheap as bottled water. Hence they should reduce the costs of superconducting systems by orders of magnitude.

Although the materials, like all ceramics, are brittle, workers have succeeded in drawing them into fine, flexible wires—a first step toward creating transmission lines, motors, generators, magnets and electrical storage devices. With techniques of particle deposition like those used to fabricate electronic chips, several groups have formed the materials into thin films. Among other applications, such films could serve as connectors between integrated circuits, thereby making computers faster and more compact. Clearly these advances promise dramatic change that will affect a broad range of economic sectors.

Analysis of the microstructure of the new materials has offered some clues to how they may superconduct at high temperatures. Several laboratories have prepared single-phase samples of the yttrium compound and, by aiming beams of X rays, electrons and other forms of radiation through the samples, have determined the compound's molecular structure. The basic unit is a perovskite (a cubic crystal common to many minerals and synthetic compounds) formed of layers of copper oxide, barium and yttrium.

The oxygen atoms in the structure have an unusual disposition, which investigators think plays a key role in superconducting. Oxygen atoms are sometimes—but not always—missing from various positions within a unit. Consequently the subscript indicating the number of oxygen atoms in a unit is not an integer: it hovers just under 7. Another significant feature of the structure is its anisotropy: it seems to allow conduction in only two dimensions, along the copper oxide layers and not in the intervening layers of yttrium or barium oxide.



OXYGEN-DEFECTIVE PEROVSKITE forms the crystalline unit of yttrium–barium–copper oxide, a ceramic compound that superconducts at above 90 degrees Kelvin, according to neutron-scattering tests done at the Argonne National Laboratory. The oxygen in two copper oxide sheets at the center of the unit buckles toward an yttrium atom. Since oxygen is missing from two sides of the "squares" at each end of the unit, the copper and oxygen there form one-dimensional chains. Oxygen may be absent in other positions.

Theorists are struggling to relate these findings to long held notions of how superconductors work. The standard model is the BCS theory, named for the three Nobel laureates who proposed it 30 years ago: John Bardeen, Leon N. Cooper and J. Robert Schrieffer. The theory postulates that superconducting occurs when electrons become bound into pairs, called Cooper pairs; pairing greatly reduces the electrons' susceptibility to thermal scattering. It was widely believed this pairing could occur only at very low temperatures. The pairing mechanism in the low-temperature superconductors that existed then was determined to be acoustic vibrations, known as phonons, in the crystal lattice. Phonons are often, but erroneously, considered to be integral to the BCS theory; actually the theory does not specify how pairing occurs.

So far there has been no serious challenge to the theory's basic premise. Theorists have directed their efforts, rather, either at explaining how phonons might cause coupling at 90 degrees K. or at proposing an alternative mechanism. The problem with phonons is that in low-temperature superconductors they provide a relatively low-energy bond that is easily disrupted by thermal vibrations.

Indeed, recent tests carried out at the AT&T Bell Laboratories have cast doubt on the role of phonons in high-temperature superconducting. AT&T workers substituted heavy isotopes of oxygen for the normal oxygen atoms in the yttrium compound. To the degree that phonons cause superconducting in the material, the T_c should have dropped, since phonons could not vibrate the heavy oxygen isotopes as energetically. Early results, according to Bertram Batlogg of AT&T, indicate that the isotope substitution makes no difference in the T_c of the yttrium compound.

Schrieffer, who is now at the University of California at Santa Barbara, had maintained that the phonon might play a role in high-T_c superconducting "if it's pushed hard enough" by some other mechanism. He says he is delighted with the fact that the AT&T results suggest an entirely new mechanism is responsible for the high-T_c superconducting: "Mother Nature has done it again."

Schrieffer's former colleague Bardeen, who is at the University of Illinois at Urbana-Champaign, suspects that at high temperatures the coupling mechanism may be electronic, caused by what is known as an exciton. This so-called virtual particle supposedly consists of an electron combined with the positively charged hole it leaves when it migrates. The exciton theory, an outgrowth of ideas developed in the early 1960's, was refined by Bardeen and other investigators in 1973. On paper the theory does offer a plausible explanation for how high T_c's can be achieved, according to Bardeen; its weakness is that no direct evidence of the exciton has been observed.

Possibly the most original—and contentious—theorist to enter the fray is Philip W. Anderson of Princeton University. Of excitons, Anderson says: "That theory was a bummer when it was first proposed, and it still is." His own theory, put forward soon after the first of the new class of superconducting materials was discovered, proposes that electron coupling comes about not through electronic attraction but through magnetic interaction between electrons of opposite spins. "Repulsion of the electrons is the key," he says. "They hate each other instead of loving each other."

A hurdle for Anderson's theory may be its difficulty. "Anderson has good ideas," Theodore H. Geballe of Stanford says, "but they're hard to understand." Another investigator says of Anderson's proposal: "I think of it as poetry."

One thing that everyone agrees on, if only because so many claims that appeared to be far-fetched have already been borne out, is that superconducting at room temperature is possible. Rumors have abounded. A particularly intriguing report comes from Wayne State University, where workers applied a radio-frequency current to a ceramic oxide compound at 240 degrees K., only 33 degrees Celsius below zero, and found that the sample generated a direct-current voltage. This phenomenon should occur only in a superconductor. The sample did not pass other tests of superconductivity, however.

Patrick A. Lee of the Massachusetts Institute of Technology is one of the few theorists to offer a specific reason for the expected rise in T_c. He points out that in ordinary superconductors the ratio of the energy that binds electrons into pairs to the T_c has always been a constant, 3.53; as the so-called binding energy goes up, the T_c goes up. But tests on the new 90-degree superconductors indicate that the binding energy is too high for the T_c by a factor of two or more. Lee maintains that the peculiar two-dimensional behavior of the new superconductors actually constrains their full potential. If physicists can revamp the materials so that they superconduct in a more three-dimensional way, Lee says, their T_c may rise much higher.

Established 1845 SCIENTIFIC AMERICAN June 1987 Volume 256 Number 6

Drought in Africa

It is a recurrent and often devastating feature of the sub-Saharan climate. If leaders were to treat it as recurrent, they could deal with it in ways that would stabilize the region's farm production

by Michael H. Glantz

Time and again photographs from sub-Saharan Africa show starving children, emaciated animals, crowded refugee camps and dry watering holes. These disturbing scenes represent the toll of recurrent drought, which over the past 20 years has made the region a focus of global attention and generated large outpourings of humanitarian concern and assistance.

During each drought governments, international relief agencies and charitable organizations muster emergency aid programs and much is written about the misuse of land and desertification. Yet when the rains return, as they did last year, the tendency has been to treat drought as an event of the past, and concern is relaxed. The fact is, however, that drought is part of the region's climate and will occur repeatedly. It is closely related to the problem of achieving sustained and adequate agricultural production in most countries of sub-Saharan Africa, and so it should no longer be ignored in planning for development. The first step toward taking the long view of drought is to be clear about its causes. Then one can ask what its effects are and how they can be alleviated.

Searching for a single cause of African droughts is probably futile. There are many different regimes of local and regional climate, resulting from different atmospheric processes and topographic features. There are also many different societies in the region, employing different patterns of land use that require varying amounts of water resources.

On a global scale droughts are quite common. The pattern varies considerably from year to year; a year with few droughts may be followed by one with many. On a regional scale some areas have one season of rainfall and others have two. In some areas it rains in winter, in others during the summer. For example, the West African Sahel (the transition zone between the Sahara and the humid savanna to the south) has an eight-month dry season and a four-month wet season, which coincides with the Northern Hemisphere's summer. At the local level the variability of rainfall in both time and space within the rainy season can be quite high.

One of the major difficulties in dealing with drought is that it is a creeping phenomenon. Both its onset and its end are often hard to identify because it does not differ sharply from ordinary dry spells. "The first rainless day in a spell of fine weather contributes as much to the drought as the last day," Ivan R. Tannehill of the U.S. Weather Bureau once noted, "but no one knows precisely how serious it will be until the last dry day has gone and the rains have come again."

Drought means different things to different people, depending on their interest in rainfall or their need for it. The most popular conception of drought is that it is a meteorological event. Yet agricultural and hydrological droughts also occur. The terms are not synonymous.

Meteorological drought can be defined in terms of the degree of dryness (stated as a percentage reduction in relation to the long-term average annual or seasonal rainfall) and of the duration of dryness in a given region. There are scores of variations on this definition because the meaning is often specific to a region and depends on the human activity for which rainfall is being measured. A meteorological drought is sometimes difficult to identify with any degree of reliability, in part because of the nature of the phenomenon and in part because meteorological and climatological information in many African countries has been available for only a few years or is of poor quality. Moreover, information on rainfall alone is often not of immediate, direct or prime use to policy makers and agricultural planners because of other variables that can affect the usefulness of the rain that does fall—among them soil moisture, ambient temperature and rates of evaporation.

Agricultural drought results when there is not enough moisture available at the right time for the growth and maturation of crops. The timing of

LAKE CHAD, which forms part of the borders of Cameroon, Chad, Niger and Nigeria, has shrunk considerably since the 1960's because of prolonged drought in the Sahelian zone of West Africa. When the top photograph was made from a Landsat satellite in 1972, the lake encompassed some 25,000 square kilometers. When the bottom photograph was made from another Landsat satellite in 1979, after several years of below-average rainfall, the size of the lake had shrunk to less than 2,000 square kilometers. Many ancient dunes long covered by the waters of the lake are visible in the bottom view.

precipitation throughout the growing season is as important as the absolute amount per month or season because crops have varying needs for moisture as they develop. M. D. Dennett of the University of Reading and his colleagues Jeremy Elston and J. A. Rogers recently showed that the seasonal distribution of rain in the West African Sahel has changed, primarily because of a reduction of rainfall in August, which on the average is the wettest month. Such a shift is detrimental to agriculture. As a trend it can be detected only in retrospect.

Hydrologic drought results when streamflow falls below a predetermined level for a specified period of time. Most often such drought is defined as the level at which the reduction in streamflow is sufficient to significantly hinder certain human activities, such as irrigation and the generation of hydroelectric power. In West Africa the discharges of the Niger, Chari and Senegal rivers have declined sharply since the late 1960's.

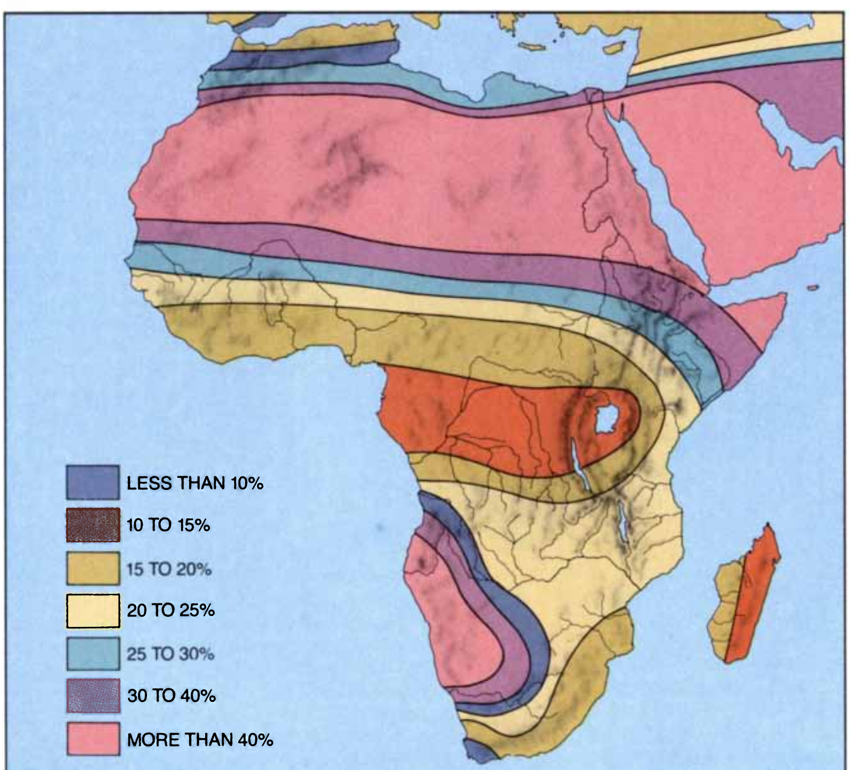
In terms of meteorological drought West Africa has had a 17-year run of "dry years." Historical records show this to be the region's third major drought of the 20th century. Looking further, one finds that the climate varies on several time scales. The three that are of interest here are the millennium, the decade and the year.

On the millennial scale several investigators cite what is known as the Milankovitch mechanism in order to explain the trend toward aridity that apparently exists in sub-Saharan Africa. In 1930 the Serbian astronomer Milutin Milankovitch proposed that changes in the elliptical orbit of the earth around the sun could affect the climate. Such changes, which occur over periods of thousands of years, are due to the gravitational pull on the earth by the major planets.

A net result was that some 10,000 years ago the Northern Hemisphere received about 8 percent more solar radiation in the summer and 8 percent less in the winter than it does now, making the summers generally warmer and the winters generally colder than they are at present. The effect of this increased seasonal amplitude in the Northern Hemisphere was to amplify both the summer and the winter monsoon circulation. It is those circulations, particularly the summer one, that account for the rainy season in the subtropics.

John E. Kutzbach of the University of Wisconsin at Madison and Alayne Street-Perrott of the University of Oxford showed that a theoretical climate model, supplied with these variations among others, is able to simulate quite well the changes in subtropical rainfall in northern Africa, southern and southeastern Asia and Central America that are evidenced by the changing levels of lakes over the past 18,000 years. The lake levels were generally highest in the period from 10,000 to 5,000 years ago; most of them have been declining since then, suggesting a very gradual trend toward aridity. Today the earth is closest to the sun during the Northern Hemisphere winter. In coming millenniums it will again be closest to the sun during the Northern Hemisphere summer. That relation should intensify monsoonal activity (and therefore rainfall) in the tropical regions.

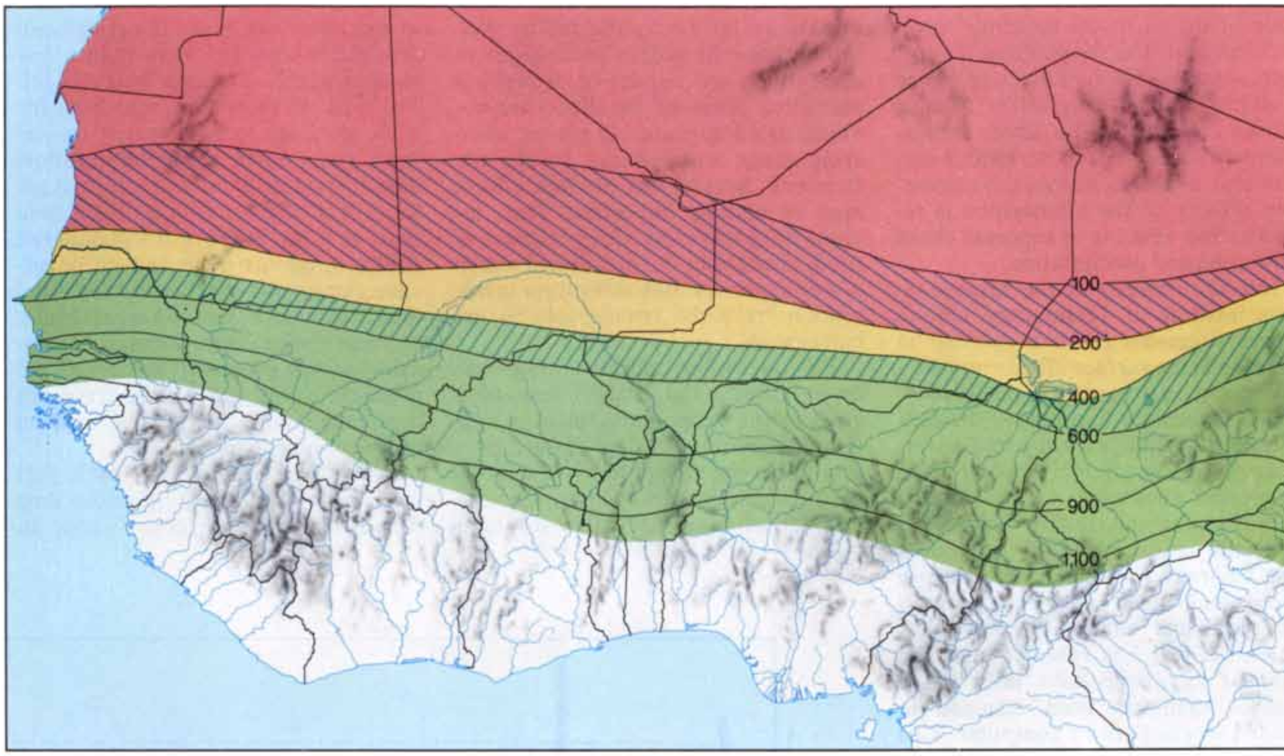
On the scale of decades and individual years the speculation about the causes of drought in Africa centers on both natural and human factors. The natural factors include random short-term fluctuations in climate, long-term climatic change, sea-surface temperature changes in the Atlantic Ocean, El Niño–Southern Oscillation (ENSO) events and ENSO-related climate anomalies, which are an example of what meteorologists call a teleconnection, meaning linkages over a great distance. The suggested human factors include both increases in atmospheric carbon dioxide and other radiatively active trace gases resulting from human activities and the modification of land surfaces.



CLIMATIC VARIABILITY in Africa is depicted in terms of the average annual departure from normal rainfall. In areas of the continent where the variability in rainfall is high and the amount of rainfall is low the likelihood of drought is a permanent feature of the climate. That is the case particularly in the regions of the continent shown in red.

Among short-term climatic fluctuations, droughts in arid and semiarid regions can be seen as part of the "normal" climate. In such areas the statistical description of average annual rainfall is skewed because a small number of years with high rainfall is averaged out by a larger number of low-rainfall years. There are not many years when the annual rainfall is near the average. To discuss drought in these regions simply in terms of departures from annual rainfall averages is therefore misleading. One must look at other statistics, such as the median rainfall (the value in the middle of an ordered range of recorded rainfalls), the range (the highest and lowest amounts) and the mode (the most frequently occurring amount) to describe adequately the rainfall characteristics of a particular African region.

Paleoenvironmental research findings show that both extended wet periods and extended dry periods have occurred in various parts of sub-Saharan Africa for thousands of years. Many investigators have searched statistical and historical records in an effort to



WEST AFRICAN REGIONS are, beginning with the red area and reading down, the Sahara zone, the Saharo-Sahelian subzone, the Sahel zone, the Sudano-Sahelian subzone and the Sudan zone. The average annual rainfall in millimeters is shown at the right.

identify drought cycles in specific regions, but their claims of having found such cycles have gained little scientific support. Indeed, they have served to suggest the absence of strong periodicities. Hence drought appears to be an aperiodic phenomenon as well as a recurrent one.

As for long-term climatic change, in the early 1970's climatologists debated whether average global temperatures were increasing or decreasing. Supporters of the cooling hypothesis contended that a new glacial period was imminent because the current interglacial period had lasted for approximately as long as interglacials usually last (between 10,000 and 15,000 years). They pointed out in addition that in only about 25,000 of the past 500,000 years was the earth as warm as it has been in the 20th century.

About 10 years ago the dominant view shifted toward the warming hypothesis. Detailed observations revealed that the cooling trend in the Northern Hemisphere beginning in about 1940 had reversed by the mid-1970's. The longer global trend toward warming, which began in about 1900, has been attributed mainly to the increase in atmospheric carbon dioxide. It is not clear whether a global warming will bring more or less rainfall to the African regions that are currently viewed as drought-prone.

El Niño is the temporary invasion of warm sea-surface water into the eastern equatorial Pacific off the coasts of Peru and Ecuador. It is a local manifestation of the phenomenon called the Southern Oscillation, which is the seesawing of mean pressure differences (at sea level) between the western and the eastern equatorial Pacific. ENSO events have been associated with droughts and other anomalies of climate around the world. The ENSO event of 1982–83 was the most intense one in at least a century with respect to the large rise in sea-surface temperature, the geographic scope and the societal impacts. Eugene M. Rasmusson of the University of Maryland at College Park sees a strong correlation between ENSO events and rainfall in southeastern Africa (the region of Mozambique and Zimbabwe), where he has found that 22 of the past 28 events have coincided with reduced rainfall. He finds much weaker correlations, however, between ENSO events and rainfall in Ethiopia, the West African Sahel and East Africa (the region of Tanzania, Kenya and Uganda).

West African droughts may be better explained by sea-surface temperature variations in the Atlantic. Climate-modeling work by D. E. Parker, C. K. Folland and T. N. Palmer of the British Meteorological Office supports the view that, as they put it, "warmer than normal waters in the tropical south Atlantic, especially in the Gulf of Guinea, have tended to favour dry conditions in the Sahel wet season, as a result of changes in the atmospheric circulation and moisture transport in the Tropics."

The principal human activity of concern is the burning of fossil fuels in historically unparalleled amounts. A growing number of scientists support the view that the increased loading of the atmosphere with carbon dioxide and other radiatively active gases such as methane, ozone, fluorocarbons and oxides of nitrogen is heating up the lower atmosphere. They are the "greenhouse gases," meaning they are transparent to shortwave (visible) solar radiation but absorb or radiate back to the earth the long-wave (infrared) solar radiation returning to the atmosphere from the earth's surface.

A warming of the lower atmosphere will affect hydrologic processes and the location of rainfall regimes, although the regional effects are not yet well understood. Nevertheless, some atmospheric scientists have speculated that the recent prolonged drought in Africa may be a first manifestation of the regional impact of such a warming.

The second human activity of concern is the modification of land surfaces by deforestation, overgrazing and desertification as well as woodcutting for fuel and construction. Such activities can increase the surface albedo of the earth—its tendency to reflect sunlight. The result is that the surface absorbs less sunlight and hence becomes cooler. This effect in turn causes changes in the lower atmosphere. With cool air at the earth's surface and warm air above, the convective activity in the atmosphere is reduced. The effect is to suppress cloud formation and precipitation.

Jule G. Charney of the Massachusetts Institute of Technology, among others, suggested a decade ago that an increase in the surface albedo reinforces drought on a regional scale. According to Charney's hypothesis, drought becomes self-perpetuating and severer as progressively more people have to support themselves on a dwindling land resource and thereby further denude the land surface of vegetation. The Charney hypothesis has recently been challenged. Historical research on ecological changes in the West African Sahel has revealed that the changes in albedo have been much smaller than has been assumed in models worked out by computer. This finding means that although albedo changes may be appearing, they are not likely to have a pronounced regional effect.

It has also been suggested that changes in the land surface might affect rainfall in other ways. For example, they could reduce the number of ice nuclei in the atmosphere that result from the decomposition of leaves and other vegetative debris. Such nuclei help to initiate precipitation. Research has suggested that the organic nuclei are better rainmakers than inorganic ones such as dust because the inorganic nuclei need much lower cloud temperatures to initiate freezing.

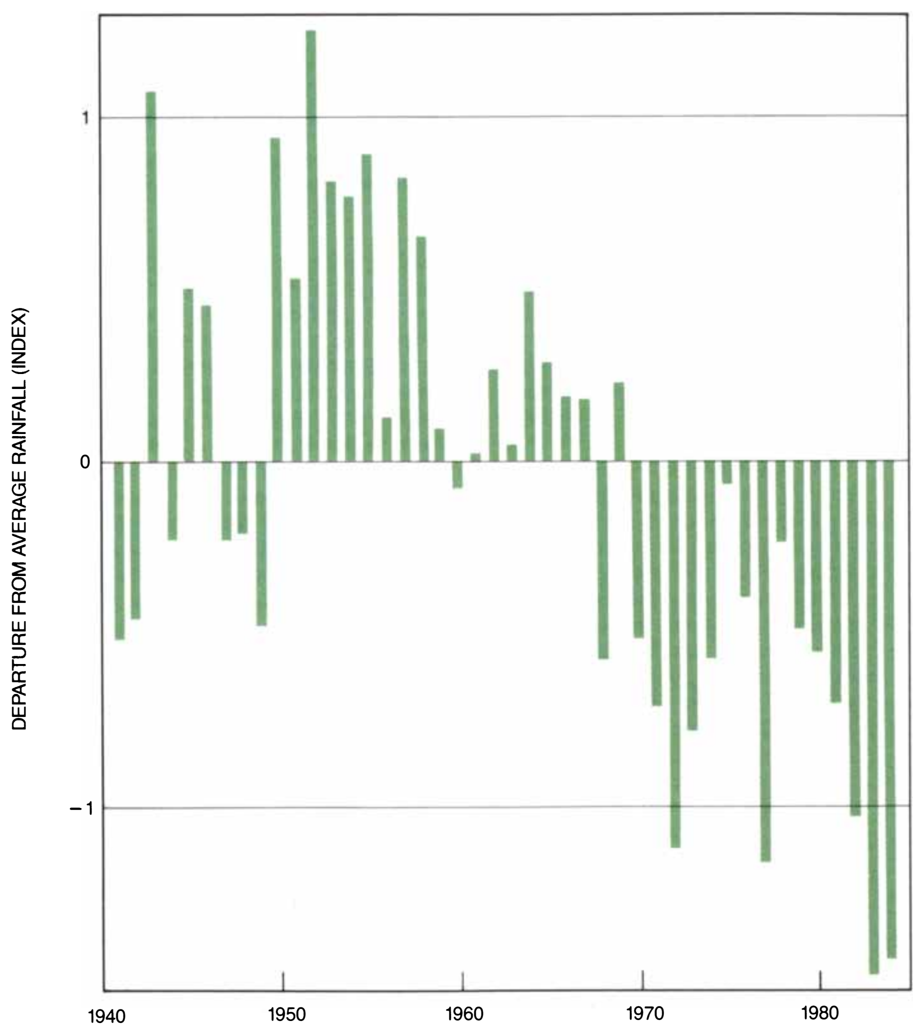
Desertification is another change in the land's surface that might affect rainfall, because it can increase the amount of dust in the lower atmosphere. Such dust absorbs and scatters sunlight, warming the upper part of the dust layer and preventing some of the solar radiation from reaching the relatively cooler surface of the earth. Again the atmospheric activity that leads to rainfall is reduced.

Studies of these geophysical conditions deal with only one set of causes of prolonged African droughts and famines. They neglect an important component of the picture: the complex interactions of climatic variability and human activity. These interactions must be taken into account if one is to understand how droughts affect agriculture, ecosystems and the economy.

In sub-Saharan Africa, where the welfare of more than 80 percent of the population is directly affected by rainfall because the people are engaged in agriculture, the impact of drought is pervasive. Some of the effects are obvious: dried-up watering places, withering crops and reduced forage for livestock. Several less obvious effects may be equally important. They include price increases, rising imports of food, changes in the nutritional status of populations at risk and surges in migration from the countryside to urban centers.

The migrations reflect the fact that crop failures and sharply increased grain prices in the marketplace resulting from prolonged drought are devastating at the village level. The hardest hit are the poorest peasants, whose reserves of grain are often low and whose debts are high. If agricultural drought persists for more than a few seasons, people abandon their villages. The men go away first, searching for work in order to earn wages to buy food. The women and children follow later to reunite families or just to get food. If drought continues for a long time, as it has in parts of sub-Saharan Africa, migrants often end up in refugee camps in a weakened condition, totally dependent on food relief. Many of these dislocations become permanent, leaving progressively less of the population engaged in agriculture and thus accelerating the decline in food production per capita.

Pastoralists are often the first to feel the impact of a drought because they usually live along the desert's edge. In



RAINFALL INDEX (*left*) for a region of West Africa including the Sahel was constructed by Peter J. Lamb of the Illinois State Water Survey on the basis of data from 20 rainfall stations. It shows below-average rainfall from 1968 through 1985. The rains improved

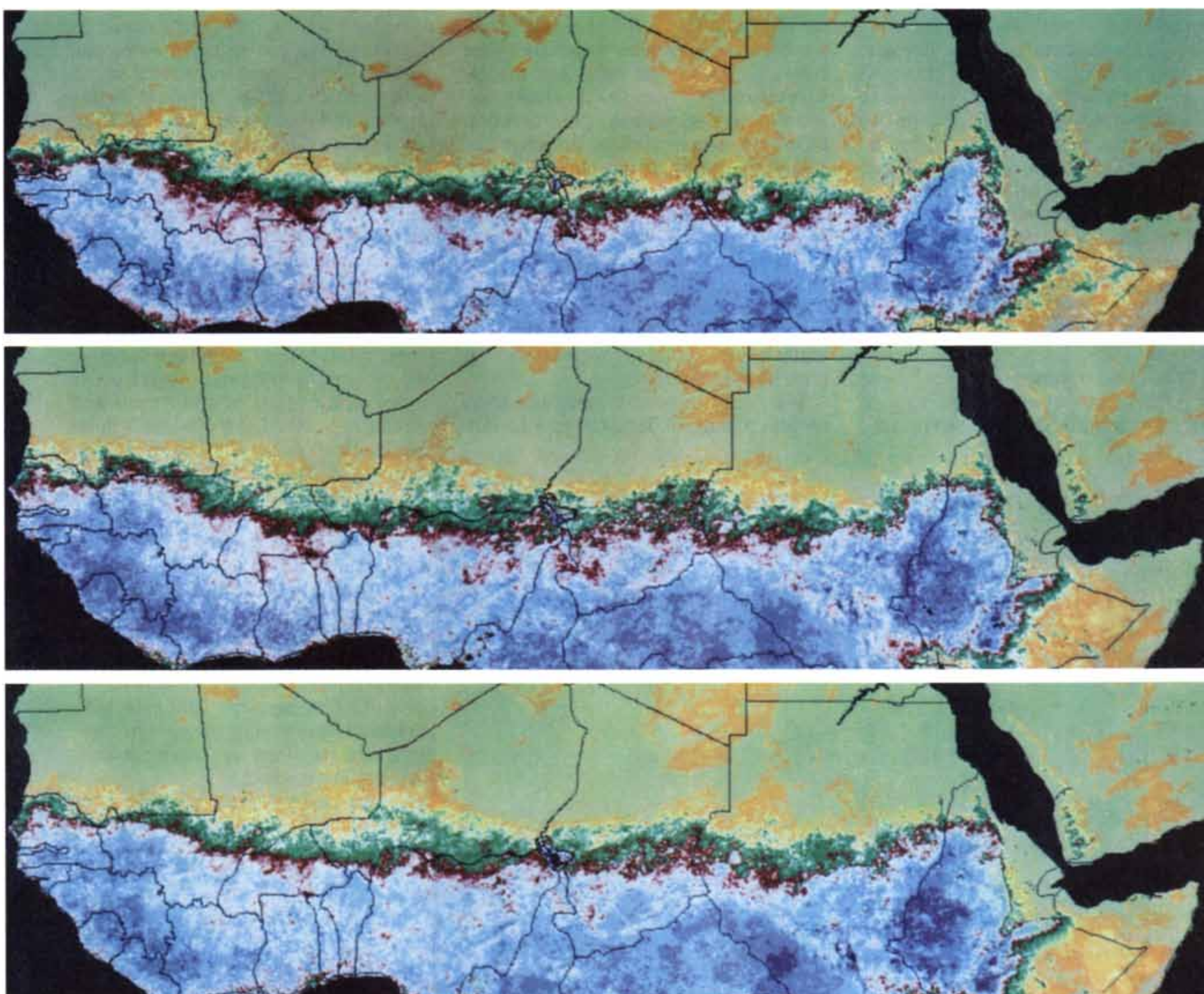
the Sahel, as the rains fail to move as far north as usual, the rangelands deteriorate and there is less forage for the herds. The pastoralists are forced to find new areas for grazing their livestock. In extreme drought conditions many animals die, as much for lack of vegetation as for lack of water. This condition often arises because livestock overgraze the vegetation around permanent and semipermanent watering points and the balance between vegetation and water resources is destroyed. Many pastoralists end up in refugee camps along with the poor farmers, or in urban centers where they become accustomed to urban life and lose interest in returning to the rangelands and the struggle with the vagaries of climate.

In the urban centers and refugee camps they and other refugees acquire a taste for imported grain such as wheat and rice instead of the traditional crops such as millet and sorghum. Hence a long-term and sometimes hidden effect of drought is an increased demand for food-grain imports, which causes a heavy drain on a developing country's precarious reserves of foreign exchange.

To support such costs the governments must devise ways to generate foreign exchange. One of the major ways to do so is to grow cash crops such as cotton, peanuts and coffee for export. These ventures take place at the expense of the traditional food crops, inasmuch as the cash crops are usually grown in areas that have good soil and adequate water supplies and that previously supported food-crop production. It is noteworthy that even during the severe droughts in the West African Sahel and Ethiopia in the early 1970's and 1980's the production and export of cash crops were maintained and even increased at the same time that the production of food crops declined sharply.

Droughts occur in many parts of the world every year, but they do not necessarily result in famines or severe food shortages. Brazil, India, Indonesia and Kenya are examples of nations that have coped with difficult drought conditions in recent years by a variety of mechanisms.

Only a few of the African countries



somewhat last year, and the resulting effect on vegetation can be discerned in satellite data (*right*) assembled by the Food and Agriculture Organization and the National Aeronautics and Space Administration. The satellite photographs, from the top down, reflect the situation in August and September of 1984, 1985 and 1986. Light green represents the least vegetation, dark blue the most.

that registered food shortages in 1982–84 actually suffered famine. The countries that did have famine (Mozambique, Angola, Sudan, Chad and Ethiopia) were plagued not only by severe drought but also by internal war. It is evident from this relation that drought by itself does not determine whether there will be a famine. It does, however, heighten other problems within a society.

It is possible to improve the reliability of projections about a society's vulnerability to drought by employing historical information on similar situations and by studying analogous experiences in other places. There is a high probability that a country in which there are internal conflicts is at a considerable risk of famine if there is a drought. The recent experience of Ethiopia is a case in point. Moreover, but less obviously, governments that appropriate the best land for cash crops displace local farmers and herders, forcing them to eke out a living in the relatively more marginal areas. As arid and semiarid lands less suitable to rain-supported agriculture are put to the plow, the probability that drought will affect farming adversely and lead to desertification will increase. This will be the case not because patterns of precipitation and the probability distribution of rainfall have necessarily changed but because the new activities will require long-term water supplies exceeding what the rainfall can provide.

Although irrigation has been seen by African governments and foreign donors as a buffer to drought, it often has been ineffective. Irrigation is expensive. The crops grown on irrigated land require not only the infrastructure of pumps, pipes and water channels but also such costly inputs as fertilizer, herbicides and pesticides. Traditional food crops are usually not able to bear the cost of such systems because food prices and the prices farmers get for their crops are kept artificially low by government policy in most African countries. As a result the governments typically develop irrigation schemes for cash crops such as cotton and sugar in order to generate much needed foreign exchange. Today irrigation encompasses only a small percentage of Africa's agricultural land, and it is not likely to be widely adopted for the production of traditional food crops.

Other proposals for providing a buffer against drought include various schemes to modify climate and weather. They can be classified according to whether they involve vegetation (the construction of tree belts and the revegetation of desertified areas), atmospheric circulation (the creation of inland seas in ancient drainage basins to increase the amount of moisture in the atmosphere by evaporation) or precipitation (the seeding of large-scale monsoonal fronts and local clouds). Such technological fixes based on scientific hypotheses are often of questionable value and may in fact serve only to conceal deeper ecological and social processes and to raise hopes falsely.

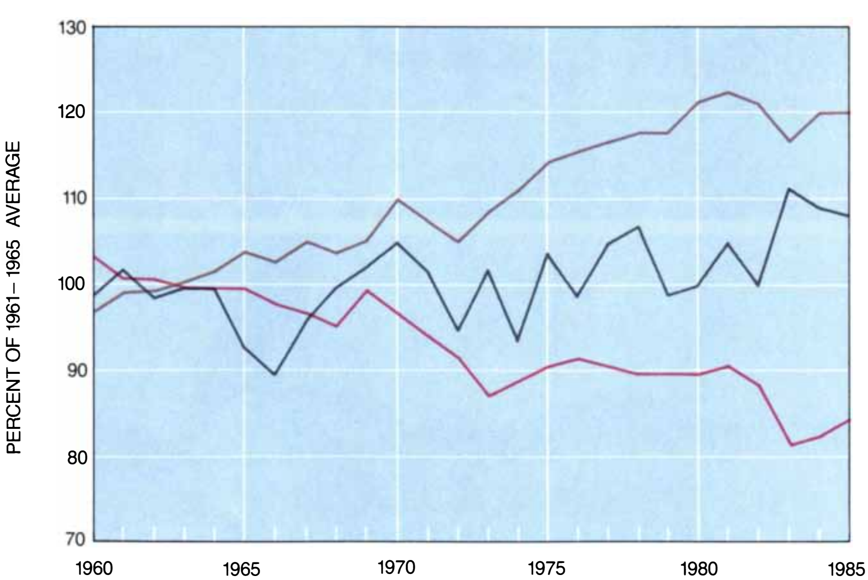
The only plausible medium-term options are those designed to alleviate the impact of drought on African societies. They include shifting from cash crops to food crops during a drought or ensuring that the foreign exchange derived from the export of cash crops goes to buy food for the affected populations; providing donor aid that is relevant to the needs of the recipients, and taking into account the agricultural and climatic situation in setting up new food-production schemes. Above all, such options must be based on an understanding that meteorological drought alone does not usually wreak the havoc often ascribed to it. The point to grasp is that the effects of drought can combine with underlying social, economic or political problems to devastate a society's food-production capability.

For the longer term there are other remedial steps. One step is a program of education and persuasion by agencies such as the World Bank on the theme that drought is a major and recurring disruptive force with which policy makers must reckon. It is also important to educate the leaders of vulnerable countries about the nature of drought as a constraint on development. Such an educational program will have to be carried on without interruption, since the tenure of leaders is often shorter than the interval between droughts.

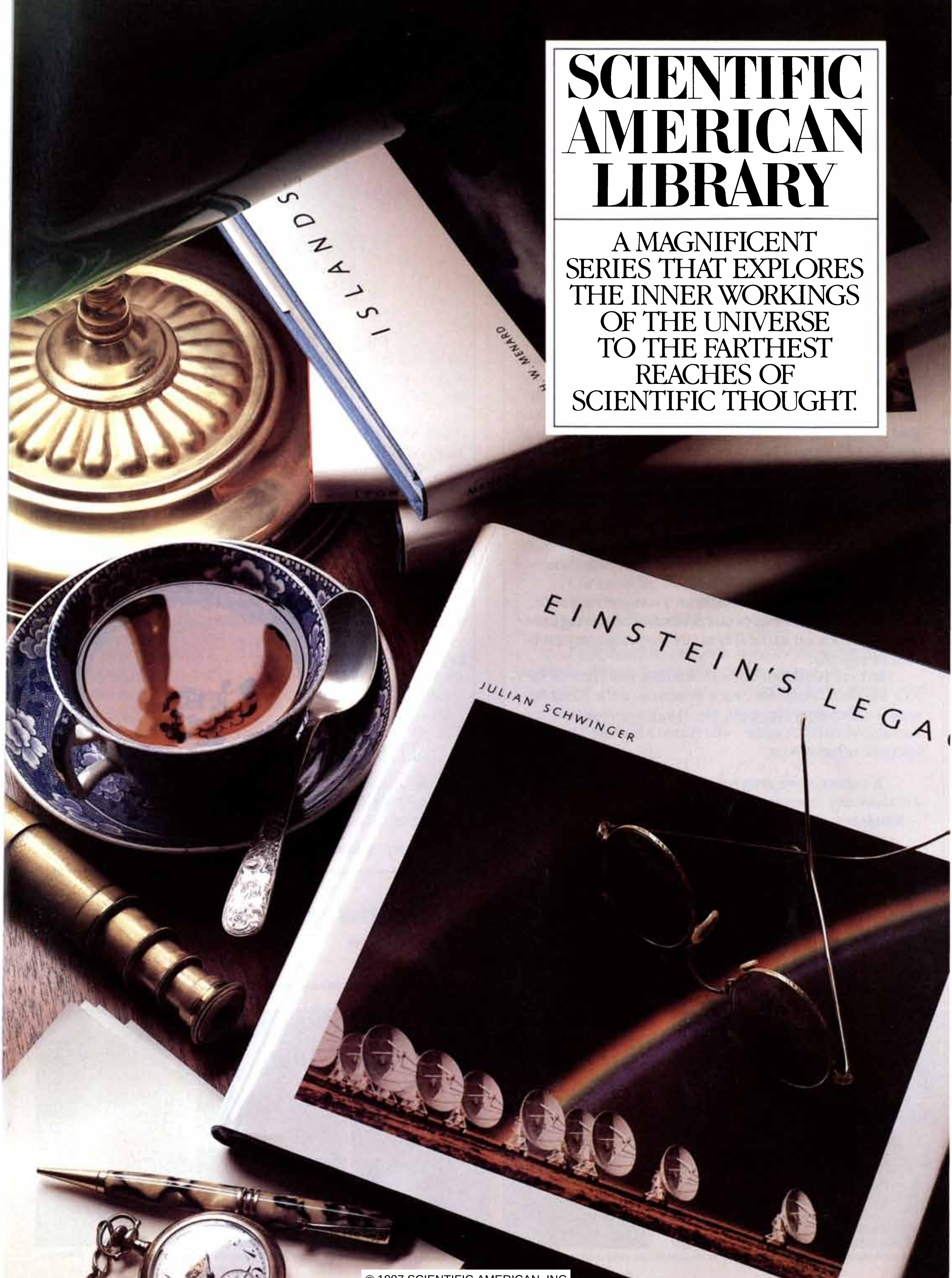
For the governments themselves it is important to adopt policies that restrain the tendency to extend farming and grazing to marginal rainfall areas or unsuitable land. When farming and grazing are extended to marginal land, the results of agricultural drought (including withered crops and dusty soil) worsen food shortages and extend desertification.

Governments could in addition make greater use of their meteorological services in agriculture. Those services can help policy makers by showing how meteorological information can improve the soundness of their decisions relating to agricultural development. The services can also take part in early-warning systems to alert governments when prefamine conditions are in the offing.

It is increasingly clear that measures of only the physical aspects of drought (magnitude, duration, intensity and geographic scope) may in fact yield little insight into why droughts with seemingly similar physical characteristics have impacts that vary from one country to another and even from one time to another in the same region. Better insights will come only from multidisciplinary studies that delve into the social, economic and cultural factors as well.



FOOD PRODUCTION per capita in sub-Saharan Africa has declined since the early 1960's, reflecting in part the scant rainfall. Sub-Saharan Africa (*red*), exclusive of South Africa, is compared with Latin America (*brown*) and six nations of southern Asia (*gray*). The data are based on information assembled by the U.S. Department of Agriculture.

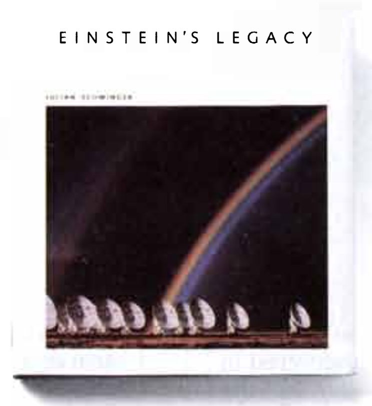


SCIENTIFIC AMERICAN LIBRARY
A MAGNIFICENT SERIES THAT EXPLORES THE INNER WORKINGS OF THE UNIVERSE TO THE FARTHEST REACHES OF SCIENTIFIC THOUGHT.
ISLANDS
H. W. MENARD
EINSTEIN'S LEGA
JULIAN SCHWINGER

Your Premier Volume

EINSTEIN'S LEGACY

Julian Schwinger



Nobel laureate Julian Schwinger tells the story of one of the twentieth century's greatest achievements—the theory of relativity—which is largely the work of one man: Albert Einstein.

The groundwork was laid in the seventeenth century by Isaac Newton, who unified motion on Earth and in the heavens in a framework of absolute space and time; then, in the nineteenth century by James Clerk Maxwell, who unified electricity, magnetism, and light. But it was left to the sixteen-year-old Einstein to glimpse, for the first time, that the theories of these two giants were incompatible.

Schwinger makes a lively narrative of Einstein's quest for the reconciliation of this conflict, a quest that led to the unification of matter and energy, and of space and time in his special and general theories of relativity.

The special theory has had its awesome confirmation in mankind's command of the nuclear force. Schwinger shows how the general theory, in turn, has stood up for 70 years to experiments drawn from the theory by Einstein, himself, and by his successors, employing ever more ingenious instrumentation and carrying the proofs ever farther beyond the decimal point.

In the last chapter, Schwinger describes space-age experiments made possible by the technologies that incorporate the theory itself. Their outcomes may not only secure further confirmation of Einstein's legacy, but face it with difficulties that lead on towards a still more comprehensive theory.

Julian Schwinger was awarded the Einstein Prize in 1951, the National Medal of Science in 1964, and the Nobel Prize for physics in 1965.

He is currently University Professor of the University of California, Los Angeles. He received his Ph.D. from Columbia University and has been on the faculty at Purdue University and Harvard University. Through the years, he has done theoretical work in various areas of both classical and quantum physics.

"I am happy to report, that his book makes for delightful and instructive reading. Technical demands on the reader do not exceed the most elementary algebra.

"It is particularly gratifying that the subject's intricacies, including glimpses of non-Euclidean and Riemannian geometries, and the theory's early successes, . . . are so well conveyed here in simple language. Altogether, this well printed and pleasingly illustrated book is an ideal gift for the curious non-expert." —NATURE

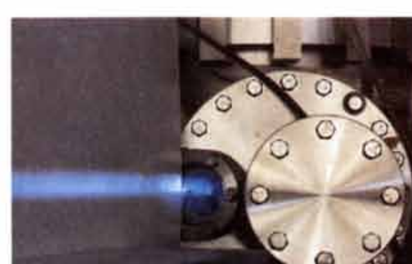
250 PAGES, 200 ILLUSTRATIONS

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

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

Radio telescopes
Stephen Northup/Black Star © 1980

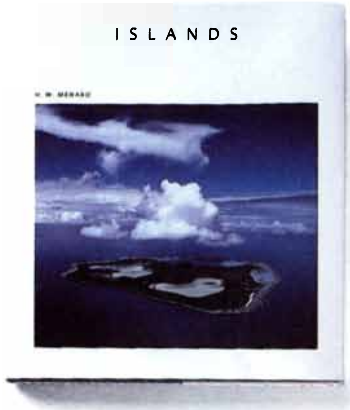
Einstein and Chaplin in Hollywood, 1931
The Granger Collection



Synchrotron light produced at Brookhaven National Laboratory

ISLANDS

H.W. Menard



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

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

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

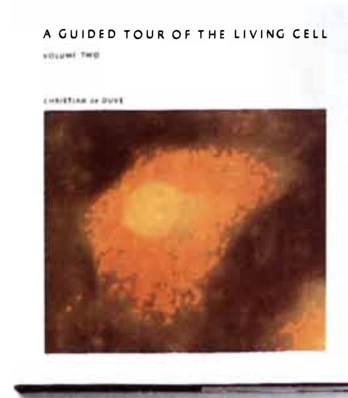
H.W. Menard was a founding member of the Scripps Institution of Oceanography, where he was professor of Geology. In the course of his distinguished career, he served as Technical Advisor in the President's Office of Science and Technology and as Director of the U.S. Geological Survey.

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

230 PAGES, 152 MAPS, PHOTOGRAPHS AND ILLUSTRATIONS

THE LIVING CELL

Christian de Duve



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

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

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

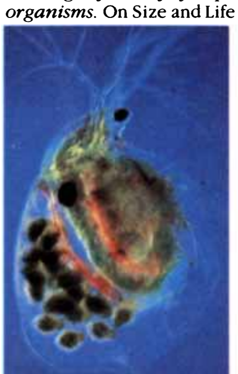
"It is hard to imagine a better introduction to the world of the cell and all it means to everyone."

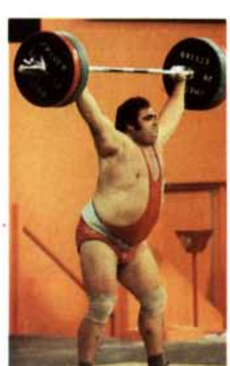
—JOHN BARKHAM REVIEWS

2 VOLUMES, 422 PAGES, 425 ILLUSTRATIONS

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

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





A molecular view of the plasma membrane. Electro through an axopod of the heliozoon Echinosphaer

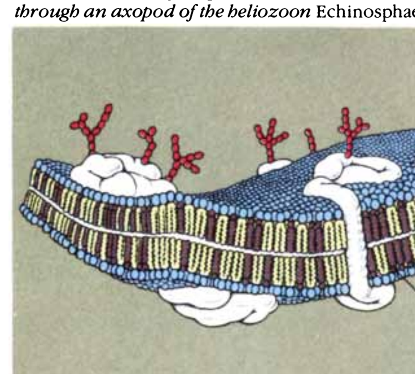
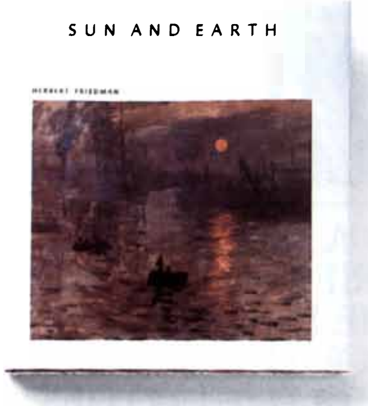


PHOTO CREDITS—Daphnia: James Bell, Photo Researchers, Inc./Ant: Daniel Adams/Elephant: Peter Johnson, National History Photographic Agency/Illustration drawn by Neil O. Hardy/Electron microgr Mutual Research Corporation

SUN AND EARTH

Herbert Friedman



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

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

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

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

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

251 PAGES, 195 ILLUSTRATIONS

OTHER VOLUMES IN THE LIBRARY:

SUBATOMIC PARTICLES
Steven Weinberg

MATHEMATICS AND OPTIMAL FORM
Stefan Hildebrandt and Anthony J. Tromba

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

HUMAN DIVERSITY
Richard Lewontin

PERCEPTION
Irvin Rock

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

THE SECOND LAW
P.W. Atkins

THE SOLAR SYSTEM
Roman Smoluchowski

THE SCIENCE OF MUSICAL SOUND
John R. Pierce

FOSSILS
George Gaylord Simpson

CONSTRUCTING THE UNIVERSE
David Layzer

FIRE
John W. Lyons

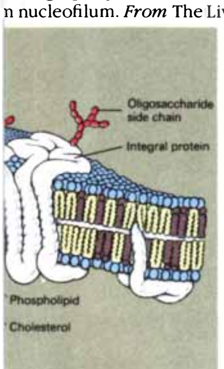
ON SIZE AND LIFE
Thomas A. McMahon and John Tyler Bonner

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

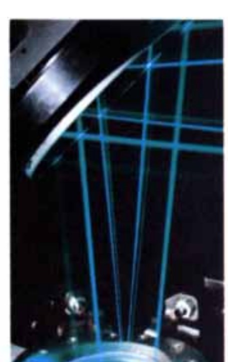
EXTINCTION
Steven M. Stanley (in preparation)

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

nicrograph of a section
n nucleofilum. *From* The Living Cell.



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



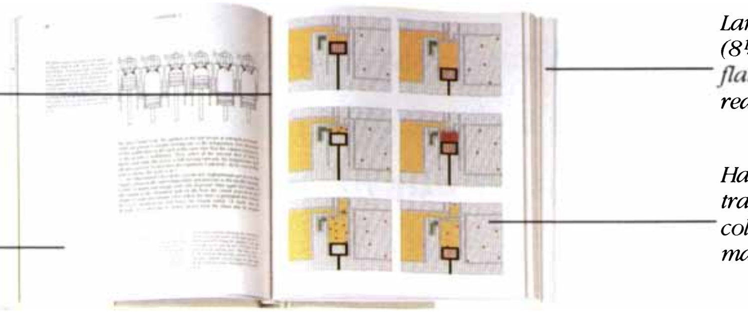
ı courtesy of Dr. L.E. Roth and Dr. Y. Shigenaka/Fire in trees: L. Burr, Gama-Liaison/Copper Smelter: John Moss, Photo Researchers, Inc./Laser: Courtesy of Sandia Labs/Plastics Fire Test: Courtesy of Factory

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

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

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

Ample margins for visual relief and notes.



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

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

JOIN OUR LIBRARY. THERE IS NO RISK INVOLVED.

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

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

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

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

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

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

MEMBERSHIP RESERVATION

Free with your paid order:
TELESCOPING POCKET TRAVEL ALARM.

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

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

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

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

Signature ____________________

Name ____________________

Address ____________________

City ____________ State ______ Zip ______

687PAP

SCIENTIFIC AMERICAN LIBRARY

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

Molecular Mechanisms of Photosynthesis

Spectroscopy, X-ray crystallography and molecular genetics combine to give a detailed picture of events in photosynthesis and show how particular molecules contribute to the process

by Douglas C. Youvan and Barry L. Marrs

Photosynthesis, the process by which light from the sun is converted into the energy necessary for the vital functions of living things, is the keystone of life on the earth. The energy captured by photosynthesis ultimately feeds not only the photosynthesizing organisms themselves but also the animals that feed on photosynthesizing organisms, the creatures that feed on those animals and so on down the food chain. How does photosynthesis work? At the molecular level, what are the interactions that capture the energy of sunlight and turn it into the energy of life? What is the architecture—the organization in space—of the molecules involved? And how does that particular architecture give rise to the speed and efficiency of the photosynthetic process?

These fundamental questions are now yielding answers. Through the separate efforts of workers in such disparate fields as spectroscopy, X-ray crystallography and molecular genetics, the molecular mechanisms of one kind of photosynthesis—the photosynthesis of certain bacteria—have been pictured in great detail.

Each area of research sheds light on a different aspect of the process. Spectroscopists have determined the sequence and timing of molecular events in the so-called light reactions of photosynthesis (the reactions constituting the first stage of the process), as well as the speed with which each interaction proceeds. X-ray crystallographers have visualized the spatial structure of the photosynthetic reaction center (the site where the light reactions take place) and have discovered how the molecules in the reaction center are aligned with respect to one another. Molecular geneticists have located and analyzed—and can now manipulate—the genes that direct the construction of the reaction center's major components.

Knowledge of the molecular interactions, structure and genetic basis of the photosynthetic reaction center makes it possible to ask more detailed questions about its function. How does the presence of each molecule in the reaction center contribute to the function of the whole? Why are certain stages of the light reactions faster or slower than others? How would the process be changed if a particular element had a slightly different shape or composition than it does?

We and others have been trying to answer such questions by applying the powerful tools of molecular genetics. By altering the genetic information that encodes certain elements of the reaction center, one of us (Youvan) has genetically engineered photosynthetic bacteria to produce reaction centers that differ in specified ways from those of unaltered organisms. By testing the photosynthetic function of the altered organisms, it is possible to learn what effect the change has had on the ability of the organism to photosynthesize efficiently.

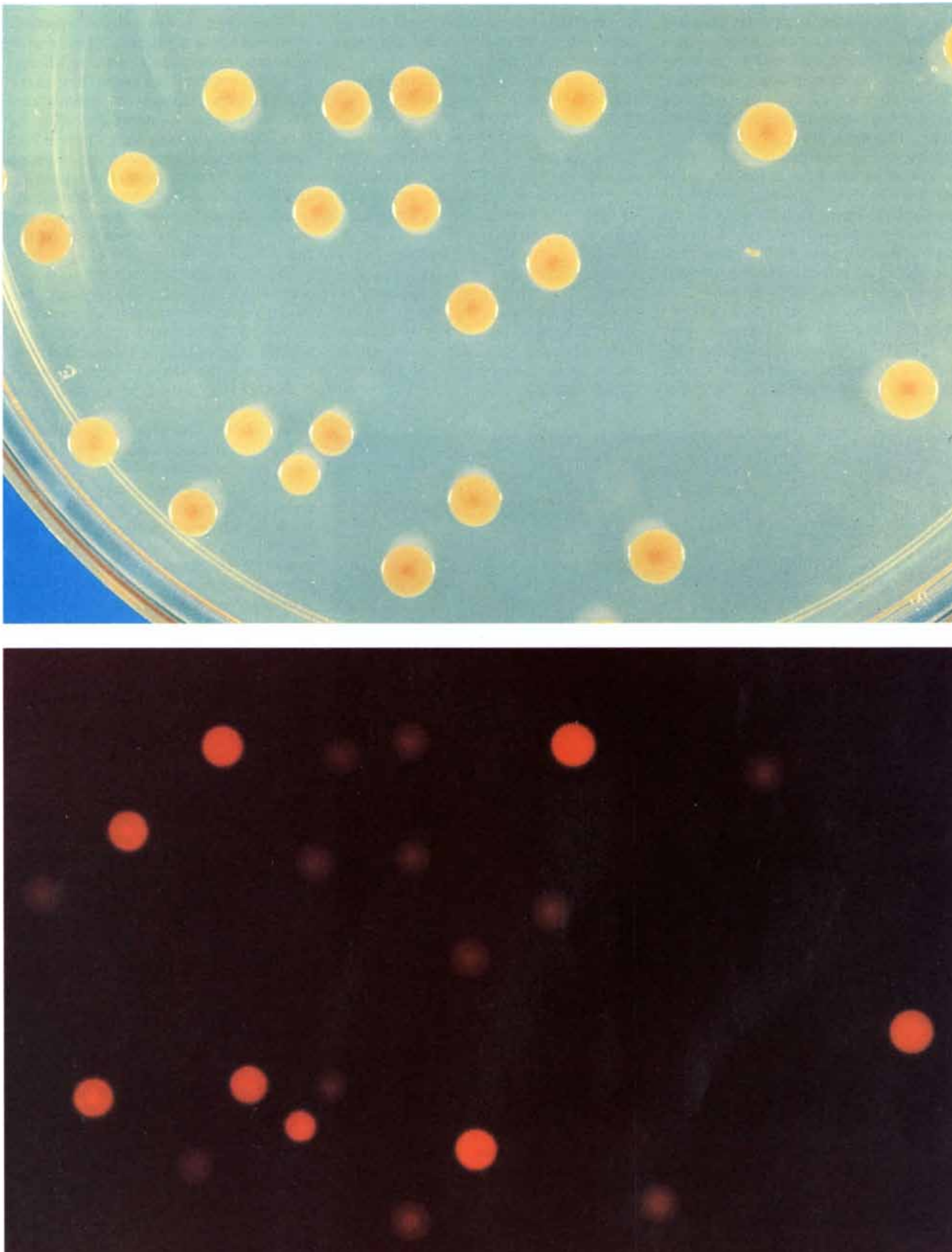
Such techniques enable researchers to learn about the mechanisms of photosynthesis in minute detail. The results—intriguing in themselves—may also make it possible eventually to design and produce organisms that can photosynthesize in specialized ways or in harsh environments.

The subjects of these studies are bacteria belonging to the genus *Rhodopseudomonas.* The rhodopseudomonads carry out photosynthesis in much the same way as the bacteria that first evolved a method of harvesting the sun's energy more than three billion years ago. The photosynthesis carried out by rhodopseudomonads differs in certain respects from the photosynthesis of higher plants; for one thing, rhodopseudomonad photosynthesis does not give off oxygen gas. There are nonetheless many similarities between the two processes. For example, both kinds of photosynthesis involve chlorophyll molecules.

A special advantage in studying rhodopseudomonads is that, unlike higher plants, they can obtain the energy required for growth not only by photosynthesis but also by several independent mechanisms, and so they can survive and multiply without photosynthesizing. It is therefore possible to grow colonies of mutants that have defects in their photosynthetic reaction centers and to study, in living organisms, the ways in which specific defects impair photosynthetic function.

The interiors of rhodopseudomonad bacteria are filled with so-called photosynthetic vesicles, which are small, hollow spheres made of lipid bilayers (the material that makes up cell membranes). The photosynthetic reaction centers, composed primarily of proteins, are embedded in the membrane of the photosynthetic vesicles. One end of the reaction center is near the outer surface of the membrane and the other end is near the inner surface.

At the start of the light reactions of photosynthesis, a photon (a packet of light energy) strikes the end of the photosynthetic reaction center nearest the inner surface of the membrane. An electron in that region of the reaction center becomes excited (elevated to a higher energy level) and carries the energy of the photon to the other end of the reaction center (the end near the outer surface of the membrane) by means of a series of chemical interactions. A second photon causes another electron to follow the same route. These interactions produce a separa-

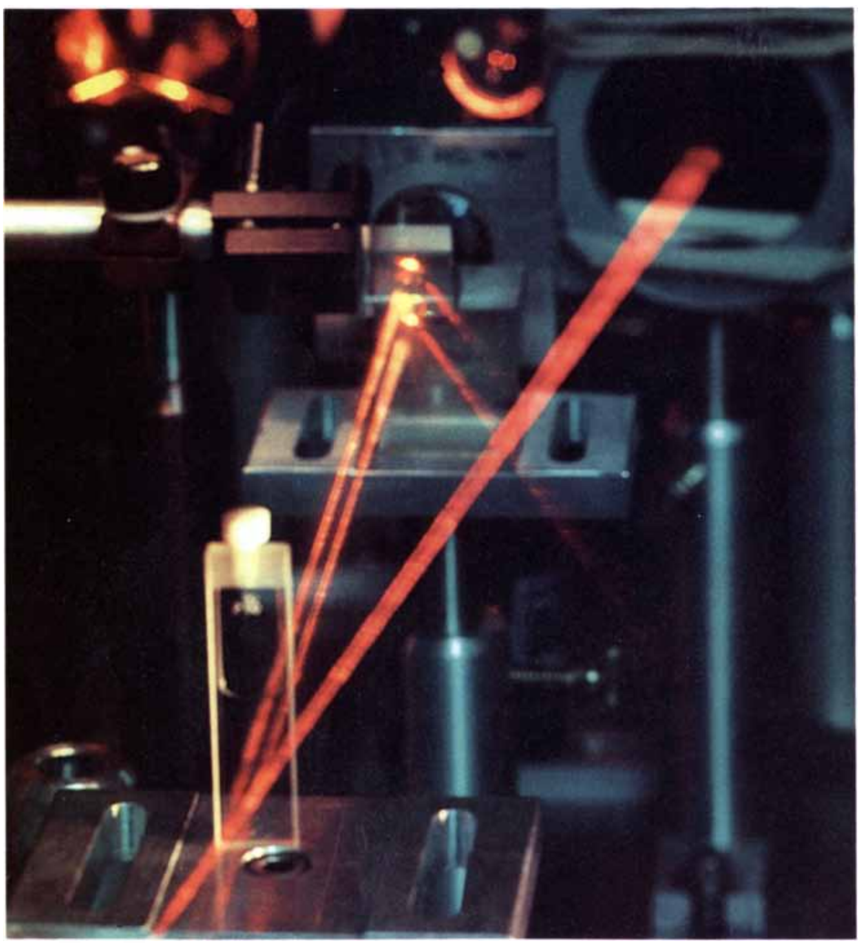


GENETIC MUTATION blocks photosynthetic function in some of these colonies of photosynthetic bacteria, which belong to the genus *Rhodopseudomonas*. Under visible light (*top*) normal (photosynthesizing) and mutated bacteria look alike. When viewed in near-infrared light, however (*bottom*), mutated bacteria fluoresce, or emit radiation, and appear bright. They are shedding light energy that they have absorbed but cannot use. The mutants survive because they can sustain themselves by means other than photosynthesis; it is therefore possible to culture mutated bacteria and study the precise effects of mutations in genes for photosynthesis.

tion of charge: the negatively charged electrons end up near the outer surface of the membrane, and molecules bearing excess positive charge, representing the absence of the electrons, are left near the inner surface of the membrane. The separation of charge represents stored energy, because energy would be released if the electrons and the positively charged molecules were able to come together. Later the charge separation drives chemical reactions that are coupled to, and provide the energy for, certain processes of the bacterial metabolism.

What are the interactions that bring the electrons from one side of the membrane to the other? That question has been answered by spectroscopists, who employ electromagnetic radiation to probe the chemical structure of matter.

One particularly useful kind of spectroscopy is called optical absorption spectroscopy. The investigator shines a beam of light through a sample and measures how much of the light is absorbed. Each kind of molecule absorbs only certain wavelengths, or colors, of light. The wavelengths absorbed depend on the structure of the molecule and its chemical environment. Every kind of molecule has its own spectroscopic "fingerprint." The spectroscopist can therefore determine the makeup of a sample by finding its absorption spectrum (the degree to which the sample absorbs light of different wavelengths).



PICOSECOND ABSORPTION SPECTROSCOPY can determine, to within a trillionth of a second, the timing and sequence of the reactions that make up the first stages of photosynthesis. Two weak probe beams (*center*) shine through a photosynthetic sample in a cuvette (*foreground*). The beams are so weak that they do not cause much photosynthesis. Photoreceptive instruments (not shown) determine the chemical makeup of the sample by analyzing the light that passes through it. An intense flash of laser light less than a trillionth of a second long (*right*) stimulates photosynthesis in a small region that is also traversed by one of the probe beams. Because the flash is so brief, all the photosynthetic reaction centers (the molecular complexes in which the first reactions of photosynthesis take place) photosynthesize simultaneously. As photosynthesis proceeds, the chemical makeup of the sample changes. The nature of the changes can be determined by analyzing the light from the probe beam. The experimental apparatus in this photograph is from the laboratory of Michael R. Wasielewski at the Argonne National Laboratory.

When the photosynthetic reaction center absorbs a photon, its chemical structure undergoes a series of changes as the interactions occur that carry the electron through the membrane. Those changes in chemistry can be monitored by noting how the reaction center's absorption spectrum changes after the center has absorbed a photon. In effect, a series of these observations traces the "chemical pathway" the electron follows from one side of the membrane to the other.

Many groups have made such observations. In their experiments the probe beam (the beam that serves to gauge the absorption spectrum of a sample of photosynthetic reaction centers, extracted from a population of bacteria) is usually kept very weak so that it will not cause much photosynthesis. Then the sample is made to photosynthesize by an intense flash of laser light less than one picosecond (a trillionth of a second) long. The laser pulse is so short that each reaction center in the sample absorbs at most a single photon, and the resulting photosynthetic reactions are synchronized: at any given time every reaction center is undergoing the same chemical transformation. Hence the absorption spectrum of the sample as a whole reflects the chemical state of each reaction center, and it is possible with successive flashes to deduce the nature and timing of the reactions in the first stages of photosynthesis.

Another method of spectroscopy is known as electron-spin resonance. Every electron has a certain amount of spin. Like all spinning charged bodies, the electron therefore creates a magnetic field; in some ways the electron can be thought of as a tiny bar magnet. In most stable compounds all the electrons form pairs in which the magnetic fields of the two electrons are aligned in exactly opposite directions. In some molecules, however, there is an unpaired electron. In electron-spin-resonance spectroscopy the investigator gauges how the energy of such unpaired electrons changes in a rapidly varying magnetic field. The results provide clues to the molecular environment of the unpaired electrons.

In the 1970's George Feher and his co-workers at the University of California at San Diego and, independently, James R. Norris and his colleagues at the Argonne National Laboratory traced the initial step of photosynthesis by electron-spin-resonance spectroscopy. They were able to see, for the first time, a signal from an unpaired electron on chlorophyll molecules that had absorbed a photon. The electron was unpaired because the chlorophyll molecules had given up an electron on absorbing the photon. The nature of the electron-spin-resonance signal indicated that the photon-absorbing element in the reaction center consists of two chlorophyll molecules in close association.

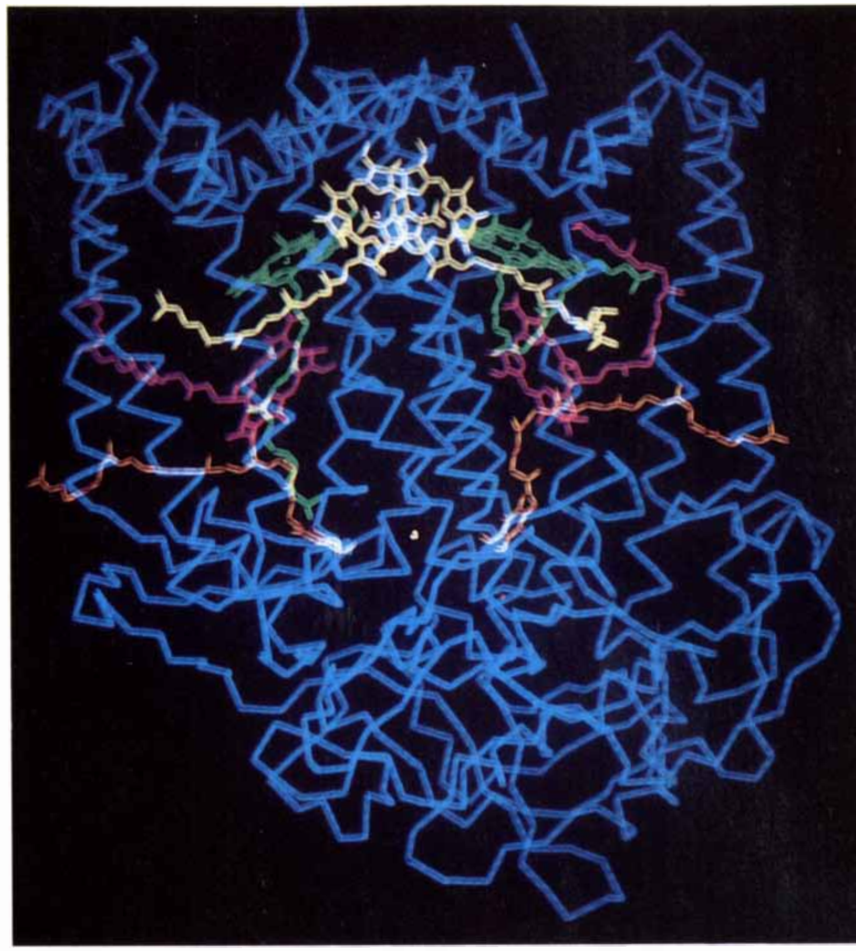
Spectroscopy has revealed the sequence and timing of the electron-transfer reactions that take place in the photosynthetic reaction center after the reaction center absorbs a photon. What about the reaction center's architecture? What physical path does the electron follow from one side of the membrane to the other?

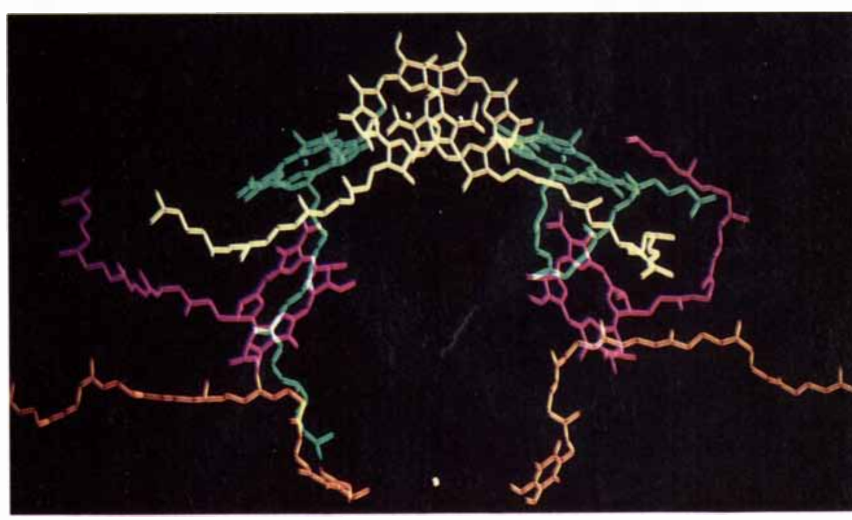
Such questions can be addressed by X-ray crystallography, which is currently the only method capable of revealing the structure of complex biological molecules at the atomic scale. The technique involves aiming a beam of X rays at a crystal and finding the directions in which the crystal diffracts the rays. From the diffraction pattern it is possible to determine the crystal's molecular structure.

The major difficulty in applying X-ray crystallography to the analysis of the photosynthetic reaction center is that the sample to be analyzed must take the form of a well-ordered crystal. Obtaining such a crystal from a protein is a difficult step, and it requires great care and some luck on the part of the crystallographer. Over the past 20 years a few hundred water-soluble proteins have been crystallized and analyzed, but it was only very recently that investigators found methods for crystallizing water-insoluble membrane proteins such as those of the photosynthetic reaction center.

It has become possible to crystallize membrane proteins largely because of technical breakthroughs in the design and use of certain small organic molecules with one end that is hydrophilic (attracted to water) and one end that is hydrophobic (repelled by water). Apparently the hydrophobic ends of these molecules can bind to the hydrophobic parts of membrane proteins, exposing the hydrophilic ends of the small molecules. The resulting conglomeration of small molecules and proteins can then be dissolved in an aqueous (water-based) solution—and so it can be crystallized. The protein molecules and the smaller molecules crystallize together, in a process known as cocrystallization.

In 1983 Hartmut Michel and Johann Deisenhofer of the Max Planck Institute for Biochemistry in Martinsreid applied such techniques to crystallize the proteins of the reaction center of rhodopseudomonads and determine their structure. The work was a tremendous achievement; the reac-



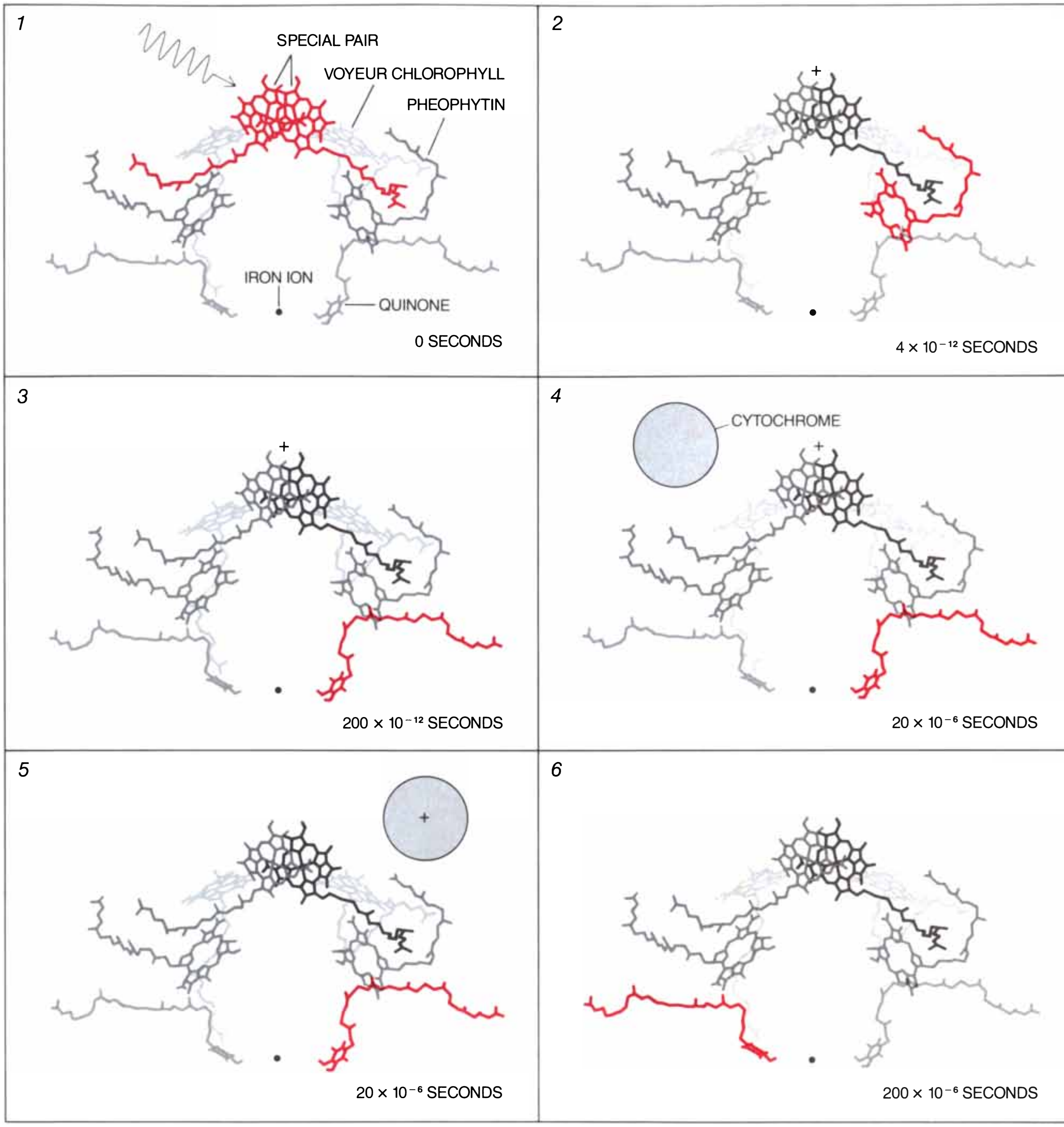


PHYSICAL STRUCTURE of the bacterial photosynthetic reaction center, shown in these computer images, was determined by X-ray crystallography. The top image shows the entire photosynthetic reaction center, the bulk of which is a protein complex (*blue*). The bottom image shows only the so-called prosthetic, or helper, molecules that are embedded in the protein. The protein itself is embedded in a membrane (not shown) that is part of an internal bacterial structure called a photosynthetic vesicle. The "special pair" of chlorophyll molecules, which absorbs the energy of photons (quanta of light energy), is shown in yellow. Other chlorophyll molecules called voyeur molecules (because they are positioned near, but may not take part in, certain reactions of photosynthesis) are shown in green. Molecules of pheophytin are shown in magenta and molecules of quinone are in orange. The small yellow dot near the base of the photosynthetic reaction center represents an ion, or charged atom, of iron. The positions and orientations shown for certain parts of the prosthetic molecules are approximate; the exact coordinates have not yet been determined. The images were produced by Chong-Huan Chang, David M. Tiede, James R. Norris, Jr., and Marianne Schiffer of the Argonne National Laboratory.

tion center is to date the only integral membrane protein whose structure is known to atomic resolution.

Together the information gleaned from such crystallographic studies and the knowledge gained by spectroscopists yield a clear picture, in space and time, of the chemical behavior of the photosynthetic reaction center and the reactions that bring the electron from one surface of the membrane to the other.

The main structural element of the reaction center is a large protein complex embedded in the membrane. A number of smaller molecules known as prosthetic, or helper, molecules (including molecules of chlorophyll) are in turn partially embedded in the protein complex. The prosthetic mole-



CHEMICAL REACTIONS that make up the early stages of photosynthesis transport an electron to one end of the reaction center while leaving behind a region of positive electric charge at the other end. In the first step of photosynthesis (*1*) a photon is absorbed by a special pair of chlorophyll molecules and transfers its energy to an electron in the special pair. (In each step the molecules bearing an electron in an excited state are shown in red.) The electron then moves (*2*) to a pheophytin molecule, passing a voyeur chlorophyll molecule and leaving a positive charge on the special pair. From there the electron travels (*3*) to the quinone molecule at the end of one spiraling chain of prosthetic molecules. At this stage a cytochrome molecule (a globular molecule) moving freely in solution approaches the special pair (*4*) and transfers an electron to it. The cytochrome molecule thereby acquires a positive charge (*5*) and the special pair is neutralized. Later the excited electron that had traveled to one quinone molecule passes to the second quinone molecule (*6*). The separation of the electron and the region of positive charge represents stored energy.

cules provide the electrically conductive path that is followed by the photoelectrons during photosynthesis.

One of the most striking features of this structure is that it has nearly perfect twofold rotational symmetry; in other words, a photosynthetic reaction center that had been rotated by 180 degrees, or half of a full circle, about its central axis (the axis running from one surface of the membrane to the other) would look almost identical with an unrotated reaction center. Specifically, the prosthetic molecules are arranged in two spirals situated symmetrically on opposite sides of the central protein. A curious and unexplained experimental observation is that, in spite of the symmetry, the electrons seem to follow only one of these spirals during photosynthesis.

The photosynthetic process begins when a photon strikes a pair of chlorophyll molecules called the special pair, situated at the end of the reaction center that is nearest the inner surface of the membrane. The special pair lies at the junction of the two spirals of prosthetic molecules. An electron within the special pair absorbs the photon's energy and moves to a neighboring prosthetic group, a molecule of pheophytin, which is very similar in composition and structure to chlorophyll. This stage of the process is extremely rapid; it takes place in about four trillionths of a second. During this reaction the electron passes near, but seems not to join, another chlorophyll molecule (known appropriately as a voyeur molecule). The special pair is left with an excess positive charge.

The electron then moves from the pheophytin to a molecule of quinone, which is at the end of the spiraling chain of prosthetic molecules and is near the outer surface of the membrane. From the quinone molecule the electron passes through the central protein to the quinone molecule at the end of the other spiral of prosthetic molecules (the spiral along which electrons are not conducted). This last step of the process is extremely slow: it occurs at a rate approximately 10^8 times slower than the electron-transfer reactions between the special pair and the pheophytin molecule.

In the meantime a globular, water-soluble molecule of cytochrome donates an electron to the special pair, thereby becoming positively charged and neutralizing the special pair. Then the entire process occurs again: another photon strikes the special pair and another electron travels along one spiral arm to a quinone molecule and through the central protein to the quinone molecule at the end of the other spiral arm.

That quinone molecule, which now carries two extra electrons, then pulls away from the protein of the photosynthetic center to participate in later stages of photosynthesis, which take place at the outer surface of the membrane. In addition a second cytochrome molecule donates an electron to the special pair, neutralizing it. The charge separation that stores the energy of the photons is now complete: two cytochrome molecules near the inner surface of the membrane have acquired a positive charge, and two electrons have traveled from the inner to the outer surface of the membrane.

The photosynthetic reaction center is remarkably effective at capturing light energy. It captures the energy of between 98 and 100 percent of the photons it absorbs. As a battery it is about 50 percent efficient: the energy stored in separated charges is about half of the energy inherent in the stimulating photons. The rest of the energy is lost in the reactions that drive the electrons along the chain of prosthetic molecules.

Spectroscopy and crystallography thus provide an intimate view of the process of photosynthesis in rhodopseudomonads. The third vantage from which this process has been examined is that of molecular genetics.

As we have mentioned, it is particularly important, from the geneticist's point of view, that rhodopseudomonads can survive without photosynthesizing. This property makes it possible for mutants that carry genetic defects affecting their photosynthetic centers to survive and multiply. Hence the effect of various mutations can be studied in live organisms.

In the genetic analysis of an organism it is important not only that mutants survive but also that there be some way to transfer DNA, the material of which genes are made, from one organism to another to produce genetic crosses. In that way altered genes can be inserted into individuals. One of us (Marrs) found about 10 years ago that a rhodopseudomonad called *R. capsulata* has just such a mechanism. It occasionally happens that the genetic material of an *R. capsulata* bacterium breaks up into sequences about five genes long. The sites of the breaks seem to be random, and they vary from one individual to another. Each piece of genetic material is packaged into a small, viruslike particle. The cell wall of the bacterium then bursts, releasing these so-called gene-transfer agents. The gene-transfer agents come to rest on other *R. capsulata* bacteria and insert their short strand of genetic material. The transferred genes are then integrated into the DNA of the recipients.

Gene-transfer agents are useful in making genetic crosses, and they can also help to "map" the genes of the bacteria. By noting how often two different genes are packaged in the same gene-transfer agent, the investigator can estimate how close together they lie on the bacterial chromosome (the strand of DNA that bears most of the bacterial genes). The more often two genes are packaged together, the closer they must be on the chromosome.

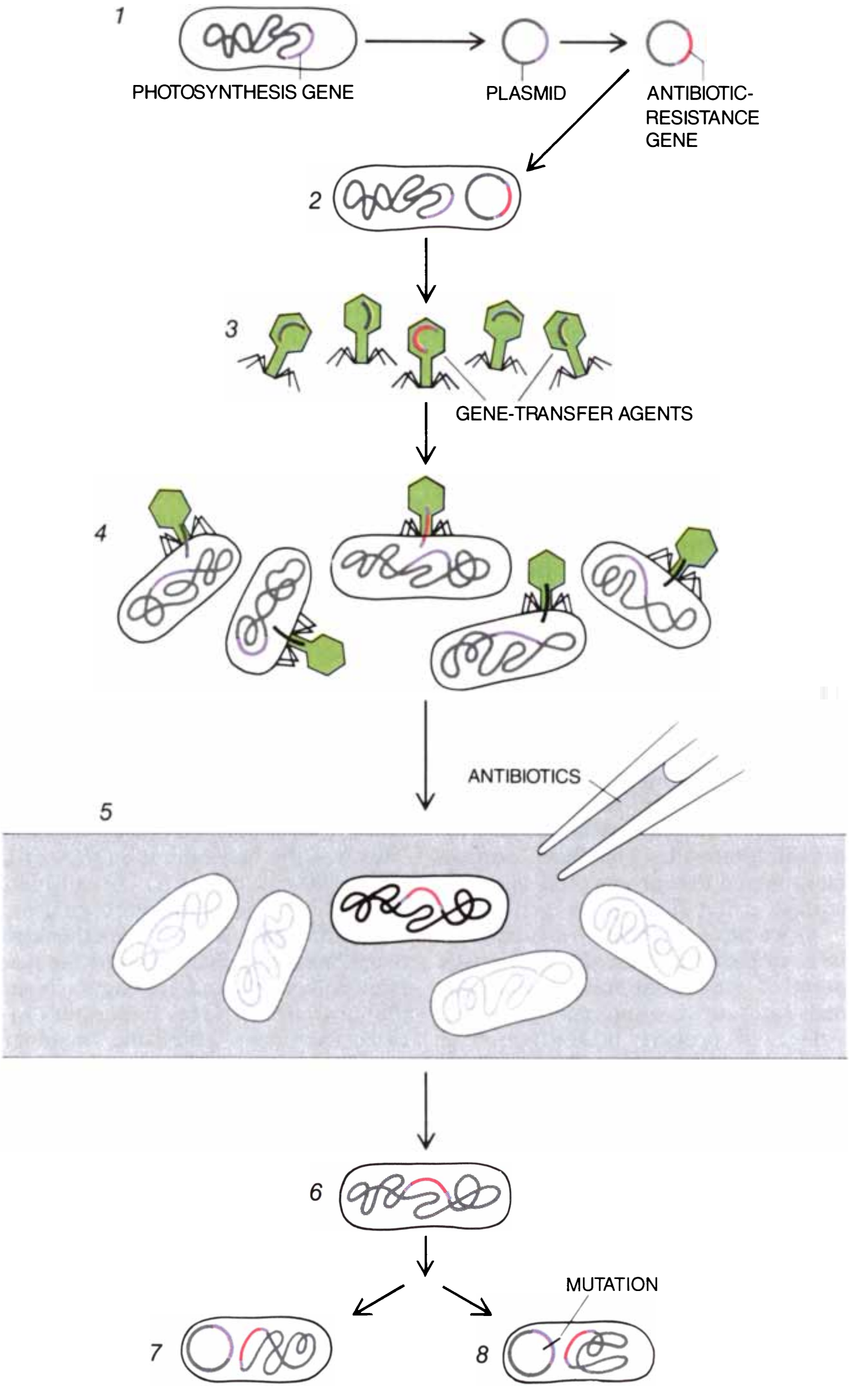
R. capsulata bacteria can also exchange genetic material by "mating." The bacteria couple by means of an appendage called a pilus, and DNA is transferred from "male" to "female" bacteria. The behavior provides another way of creating genetic crosses.

By means of such crosses, one of us (Marrs) was able in 1980 to isolate a fragment of DNA that bears most of the genes necessary for the light reactions of photosynthesis. It included genes for all the proteins within the reaction center and for proteins that make up a separate structure known as the light-harvesting antenna, which helps to capture the energy of photons and funnel it into the reaction center. This was the first time the DNA of a photosynthetic center had been isolated in any photosynthesizing organism.

Building on this work, the other one of us (Youvan) characterized the isolated DNA. He and colleagues from the Lawrence Berkeley Laboratory located each gene within the fragment and determined each gene's sequence of nucleotides (the elements that determine the specific makeup of the protein encoded by the gene). The nucleotide sequence was the basis for a computer analysis of protein structure that predicted aspects of the overall architecture of the reaction center several years before it had been visualized by X-ray crystallography. Knowing the nucleotide sequence for the photosynthetic reaction center in a bacterium also made it possible to compare that sequence with the corresponding sequence in higher plants. The close similarity between the sequences indicates that the makeup of basic photosynthetic structures is nearly universal.

Knowledge of the genetic makeup of the photosynthetic center of *R. capsulata* provides the tools necessary to address a most fundamental question: In what specific ways do the various elements of the reaction center contribute to its function?

One way to answer the question is to create simple mutations in the genes that encode the structures of the photosynthetic center. Organisms carrying



MOLECULAR-GENETIC TECHNIQUES enable an investigator to replace normal genes with mutated ones, producing a bacterium that lacks the gene for the photosynthetic reaction center, and to test the function of altered genes in the bacterium. The photosynthesis gene (*purple*) from a normal bacterium (*1*) is inserted into a plasmid (a small ring of DNA). The photosynthesis gene is then deleted, but the sections of DNA that flank it are left in place; a gene for resistance to antibiotics (*red*) is substituted. The plasmid is inserted into a normal bacterium (*2*). The bacterium breaks its DNA into small segments and packages them into viruslike entities called gene-transfer agents (*3*). The gene-transfer agents attach to other bacteria (*4*), injecting the DNA they carry. Because the antibiotic-resistance gene is sandwiched between DNA sections that usually flank the photosynthesis gene, it replaces the photosynthesis gene in the bacterial DNA. The bacteria are then exposed to antibiotics (*5*), preventing individuals without the new gene from replicating. The remaining organisms have no gene for photosynthesis (*6*). When a plasmid containing the normal photosynthesis gene is inserted into these organisms (*7*), they photosynthesize normally; such bacteria confirm that the genetic manipulations have not irreversibly damaged the organisms' ability to photosynthesize. It is also possible to insert a plasmid containing a photosynthesis gene that has a mutation (*8*). Testing the photosynthetic function of the resulting bacteria reveals the effect of the altered gene.

the mutations can then be observed spectroscopically to determine how their photosynthetic function differs from that of unmutated bacteria. Such experiments are already providing basic information on certain aspects of bacterial photosynthesis. In some cases the information gained is also applicable to higher plants.

For example, the quinone molecule that ultimately receives the electrons in rhodopseudomonads functions in much the same way, and in a similar environment, in higher plants. The X-ray structure of the protein pocket in the reaction center to which the quinone molecule binds in bacteria has recently been determined by Feher and his colleagues. One of us (Youvan), along with a graduate student (Edward J. Bylina), has introduced mutations in the quinone-binding pocket. Some of the resulting bacteria have no photosynthetic function, others have impaired photosynthetic function and still others have normal photosynthetic function; the degree to which function is affected depends on the particular mutation. In some of the mutated strains the function of the quinone molecule is changed in such a way that these strains are resistant to certain herbicides (such as atrazine) that kill plants by inhibiting the quinone molecule's function.

Such knowledge, gained from the study of bacteria, may soon be applied to higher plants. In the case of herbicide resistance, similar mutations could be made in the photosystems of such higher plants as soybeans, which are not resistant to atrazine. Soybeans made resistant to atrazine by the mutations could be planted in fields formerly planted with such crops as corn, which are naturally resistant to atrazine and are often sprayed heavily with that herbicide to kill weeds.

On a more fundamental level, the work on rhodopseudomonads may lead to discoveries concerning the basic electronic properties of various proteins. By genetic manipulations that would introduce altered reaction centers in the place of natural ones, an investigator might be able to learn how quickly or efficiently the new proteins conduct electrons. Many similar experimental tests can be envisioned.

The profound and detailed knowledge underlying such experiments has been gained by the union of three fields of research often considered to be almost completely unrelated. Each group of investigators, by examining the problem from a distinct angle, has illuminated a different aspect of the process of photosynthesis, so that it is now possible to see the whole.

DE VILLE

The Contemporary Spirit of Cadillac.

At Cadillac, we believe today's luxury cars should reflect today's values. The newly styled, longer De Ville is as contemporary as it is Cadillac. This six-passenger luxury car offers renowned comfort and convenience, enhanced by leading-edge technology. De Ville also delivers a smooth ride and precise driver control. There's even an Anti-Lock Braking System available.

The new look of luxury.

De Ville adds the distinction of a restyled front end with a new, bolder grille and composite headlamps. The profile assumes a longer silhouette. And new taillamps complete the elegant update.

The increasingly rare comfort of full-size room for six.

How can a luxury car this effortless to maneuver still accommodate six adults in Cadillac comfort? The answer is a front-wheel-drive, transverse-mounted V8 engine. By putting the engine sideways, valuable room is passed along to the interior.

Cadillac touches that anticipate your every want.

In addition to its six-way power driver seat and Electronic Climate Control, De Ville allows you to operate the stereo radio and other conveniences up to 10 minutes after you turn off the ignition, or until you open a door.

A unique blend of smooth ride and responsive handling.

With four-wheel independent suspension and Electronic Level Control, De Ville merges a silky ride with a firm feel-of-the-road. And with its available Anti-Lock Braking System, De Ville can help take the panic out of panic stops.

De Ville's new 6-year/60,000-mile warranty.

Now De Ville has a confidence-inspiring 6-year/60,000-mile warranty.*

Your Cadillac dealer invites you to experience the car that defines luxury for today. De Ville, the contemporary spirit of Cadillac.

*See your dealer for the terms and conditions of this new limited warranty.

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

Gravitational Wave Observatories

Einstein's general theory of relativity suggests that the earth is bathed with gravitational waves from distant stars. Observatories planned for the early 1990's could detect the extragalactic signals

by Andrew D. Jeffries, Peter R. Saulson, Robert E. Spero and Michael E. Zucker

Albert Einstein's general theory of relativity suggests that the earth is continually bathed with energy originating in the gravitational interactions of distant stars. According to the theory, energy released by a cosmic disturbance, such as the explosion of a star, travels outward at the speed of light in the form of gravitational waves. As the waves move, they should distort the shape of any region of space through which they pass. By measuring the motions of carefully isolated test masses it should be possible to detect the waves.

Unfortunately the perturbations of space are too weak to register with ordinary instruments, and gravitational wave detectors built and refined over the past quarter century have not yet sensed any passing waves. The difficulty of detecting gravitational waves might appear surprising given the familiarity of static gravity in such everyday experience as the falling of objects and the orbiting of planets. Yet neither of these phenomena involve masses large enough and fast enough to generate appreciable waves. The first detected signals are likely to come from distant astrophysical objects that are more massive than the sun and move at nearly the speed of light.

Plans for a new round of much more sensitive experiments to detect gravitational waves are now under way. Gravitational wave observatories currently being designed could detect extragalactic gravitational signals by the early 1990's. Prototypes of such detectors, which are based on laser interferometers, have been built and tested in laboratories both in the U.S. and in Europe. In the U.S. the four of us are working in groups led by Ronald W. P. Drever of the California Institute of Technology and by Rainer Weiss of the Massachusetts Institute of Technology to build a pair of large interferometers on opposite sides of the continent, one in the Mojave Desert in southern California and the other in Columbia, Me. The two instruments will be linked electronically to operate as a single observatory called the laser interferometer gravitational wave observatory, or LIGO. Each instrument will be 1,000 times more sensitive than the best existing detector, yet the total cost of the pair will be about that of a large optical telescope, roughly $60 million.

The view of the universe we expect the LIGO and other similar observatories to unveil will be new and qualitatively different from the universe seen through the eyes of conventional astronomy. Until the 1930's optical-frequency electromagnetic waves (visible light) were the only tool for studying the distant universe. The electromagnetic view of the universe was dramatically enhanced by the advent of radio astronomy; further advances were triggered by the opening of infrared, X-ray and gamma-ray windows. Although each form of electromagnetic energy gives a distinct view of the universe, gravitational waves carry a different kind of energy altogether, and thus a new kind of information about their sources. Gravitational wave observatories could revolutionize the view of the universe.

Demonstrating the existence of gravitational waves would also directly verify the predictions of general relativity. Of central importance in relativity is the concept that no signal can travel faster than the speed of light. In Newton's "action at a distance" theory of gravity, in contrast, changing gravitational fields would propagate at infinite speed. The simple verification that gravitational waves travel at the speed of light would lend strong support to Einstein's views.

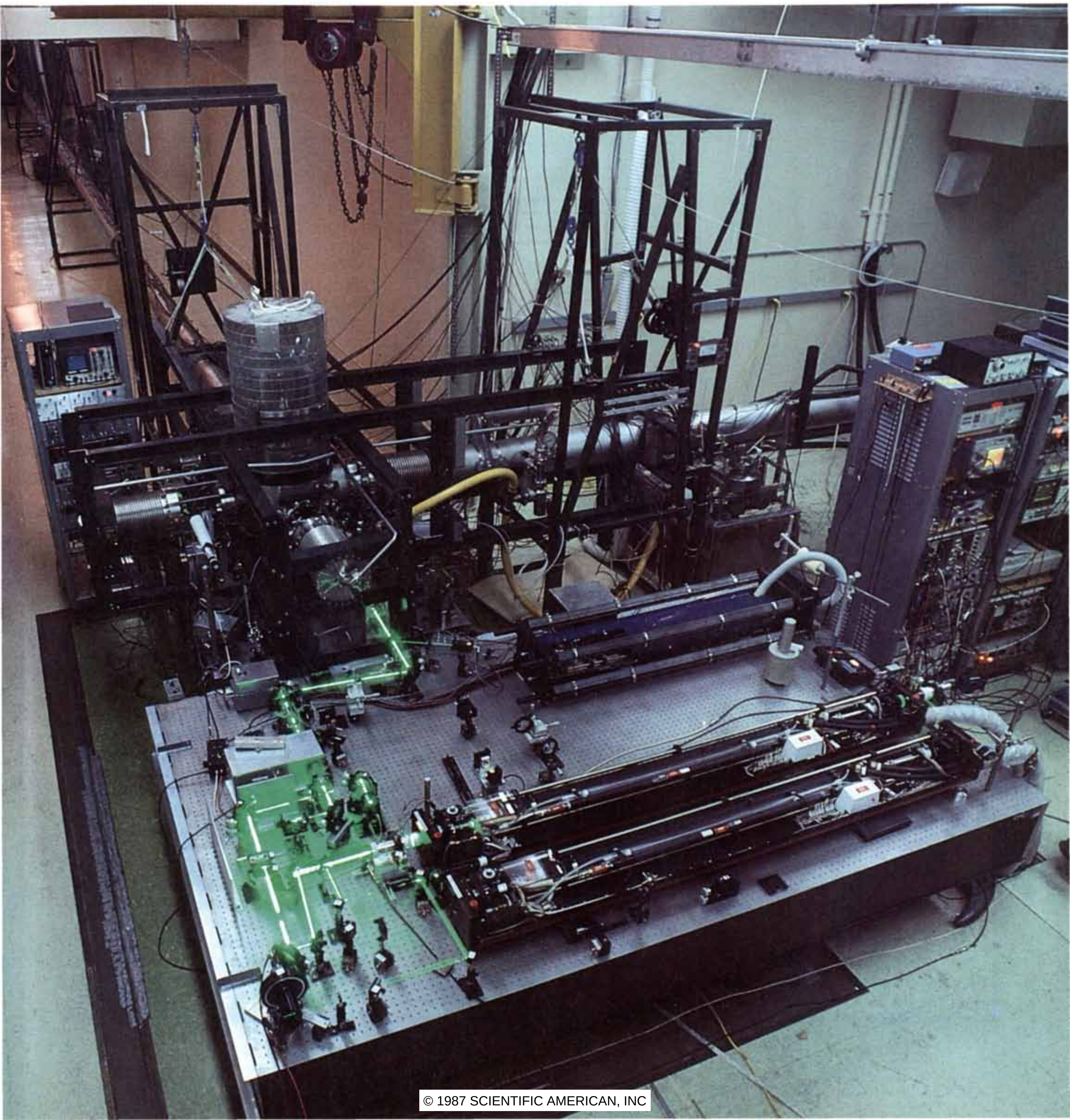
To see how gravitational waves affect matter, suppose a gravitational wave moves through space. Imagine placing a long, flexible rubber tube along the direction in which the wave travels. Initially the tube has a circular cross section. As the wave moves along the tube, it distorts the cross section into an ellipse; the major axis grows by a fraction of the original diameter, while the minor axis shrinks by the same fraction. Half a wave period later the axes interchange, so that ripples form on the surface of the tube. After the wave has passed, the tube returns to its original shape.

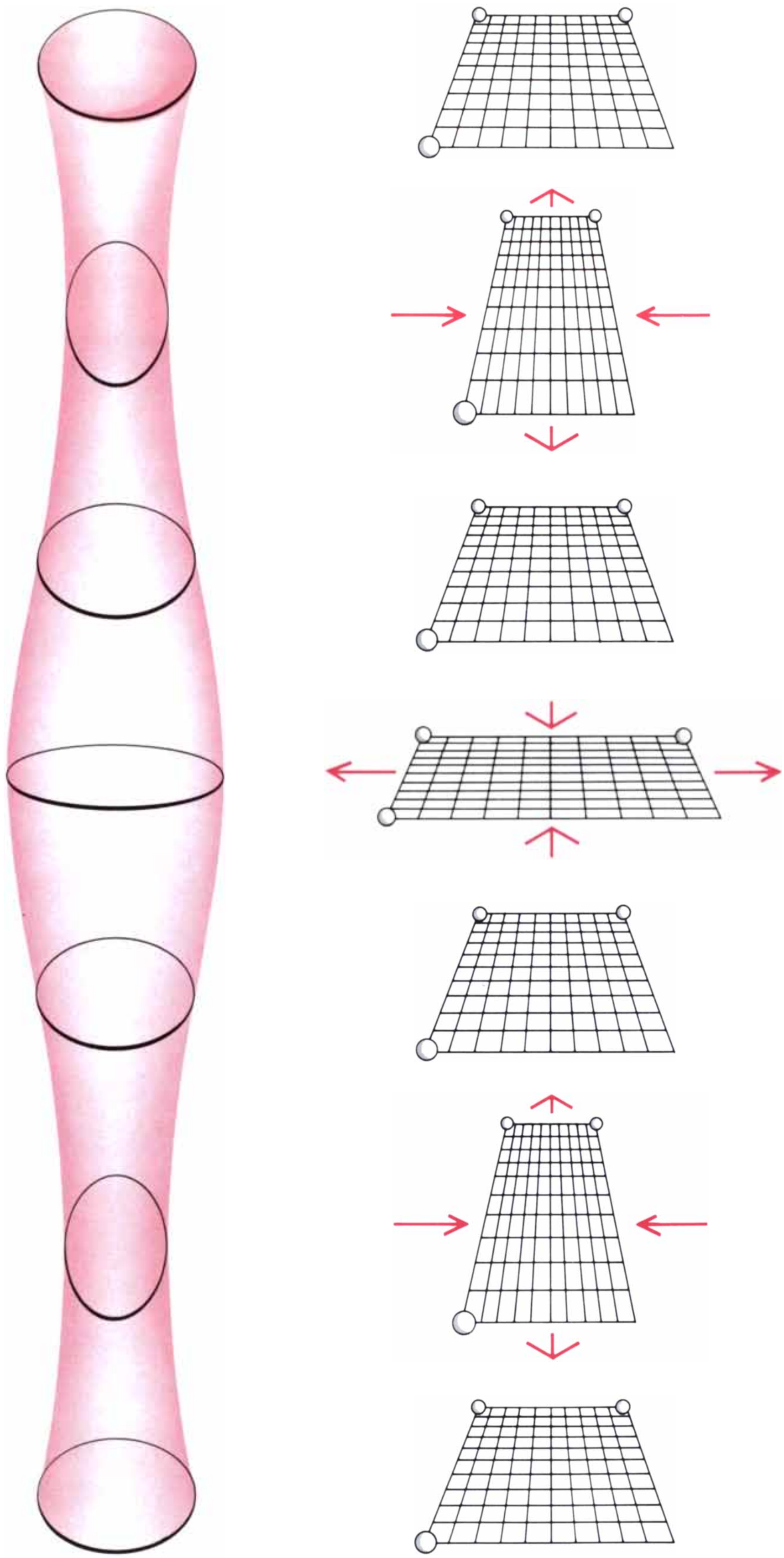
How would such a wave affect a layer of evenly spaced masses suspended in a plane? Suppose the plane is horizontal and the wave moves at right angles to it. As the wave passes through the plane, it will first increase the spac-

GRAVITATIONAL WAVE DETECTOR at the California Institute of Technology, a prototype for larger observatories planned for the next decade, consists of a laser interferometer, housed in two evacuated pipes set at right angles that stretch 40 meters down two hallways. The laser beam in the foreground is directed by mirrors and optical fibers into a vacuum tank (*left of center*). The tank contains a beam splitter, or partially reflecting mirror, that divides the light equally between the pipes. Mirrors mounted on freely suspended masses at each end of the pipes reflect the light. The light beams bounce back and forth the length of the laboratory approximately 10,000 times. The resulting interference is observed. A passing gravitational wave would slightly alter the distance between one or both pairs of masses and thereby change the interference. The apparatus is sensitive to changes as small as 3×10^{-16} meter, or one-third the diameter of a proton, lasting for as little as one millisecond. The Massachusetts Institute of Technology has a 1.5-meter interferometer and a five-meter interferometer that is currently under construction.

ing between the masses in one direction, say north-south, and decrease the spacing in the perpendicular direction, which in this case is east-west. The fractional change in separation characterizes the wave's strain amplitude, or strength. As the wave continues to move through the plane, the pattern of stretching and shrinking will reverse periodically: the spacing in the north-south and east-west directions would alternately increase and decrease. The effects induced by waves are transverse to the direction of propagation, so that a westward-traveling wave would change the north-south spacing, not the east-west spacing.

The generation of gravitational waves is similar in some ways to the generation of electromagnetic waves. Gravitational waves differ from the more familiar static gravitational attraction in the same way that light and radio waves differ from static electricity and magnetism. A moving charged object radiates electromagnetic waves with amplitude that is proportional to the electric charge and acceleration of the object. The gravitational charge of an object is its mass, and so the amplitude of a gravitational wave should be proportional to the object's mass and acceleration. But the law of conservation of momentum requires that isolated systems have no net acceleration; for each action there is an equal and opposite reaction. Most waves emitted by an accelerating mass are thus canceled by the waves emitted from





EFFECTS OF GRAVITATIONAL WAVES are depicted for a rubber tube floating in space (*left*) and for a set of masses in a plane (*right*). Gravitational waves should distort the shape of any region of space they travel through. Here the waves move vertically.

a recoiling object. The cancellation is not perfect, however, since the objects are not in precisely the same place. The slight displacement in location of the objects means that the waves arrive at different times, and so some gravitational radiation can escape.

The amount of gravitational radiation that escapes from an object depends on how unevenly the object's mass is distributed. A measure of this unevenness is a quantity known as the quadrupole moment; a soccer ball has no quadrupole moment, but an American football has a large quadrupole moment. Strong gravitational waves are emitted when a massive object undergoes rapid changes in its quadrupole moment. The mass and speed of the object must be tremendous, however; a 500-ton steel bar rotating so fast that it is almost torn apart would generate gravitational waves that would strain, or distort, a test system by at most one part in 10^{40}. Such a change is clearly too small to detect.

Stars and other astrophysical objects should emit gravitational waves that are much stronger than waves from any conceivable source on the earth. A binary star system—two stars that orbit their common center of mass—should produce continuous gravitational waves with a fundamental period equal to half of the orbital period (since the quadrupole moment takes on the same value twice in each orbit). Nearly half of all stars are paired in binary systems. Binaries have a large varying quadrupole moment, and the gravitational radiation they generate can be predicted precisely. The strongest gravitational waves emitted by binaries should induce a strain of one part in 10^{20} on the earth, which is 20 orders of magnitude greater than the strain induced by the whirling 500-ton steel bar.

A small fraction of the known binaries consist of two tightly orbiting neutron stars. A neutron star is an object, made up almost entirely of neutrons, that is as massive as the sun and yet has a radius of only about 10 kilometers. A neutron star is sometimes visible as a pulsar: a radio source that appears to flash on and off as it rotates about its axis, much like a lighthouse sweeping its beam in a circle.

In 1974 Joseph H. Taylor, Jr., and Russell A. Hulse, then at the University of Massachusetts at Amherst, discovered a pulsar in a binary system that has an orbital period of eight hours. Although only one of the stars is a pulsar, studies of the orbit reveal that its partner is also a neutron star. By timing the arrival of signals from the pulsar, Taylor and his co-workers mapped the orbit of the two neutron stars with great precision. They found that the period of the binary is gradually decreasing, which indicates that the two neutron stars are spiraling toward each other. The observed energy loss agrees accurately with the loss predicted for the system through gravitational radiation; that is, gravitational waves would carry away just the right amount of energy and angular momentum to cause the observed effect. Actually the theoretical prediction matches the measurement to within 1 percent. The impressive agreement between theory and experiment currently is strong evidence for the existence of gravitational radiation.

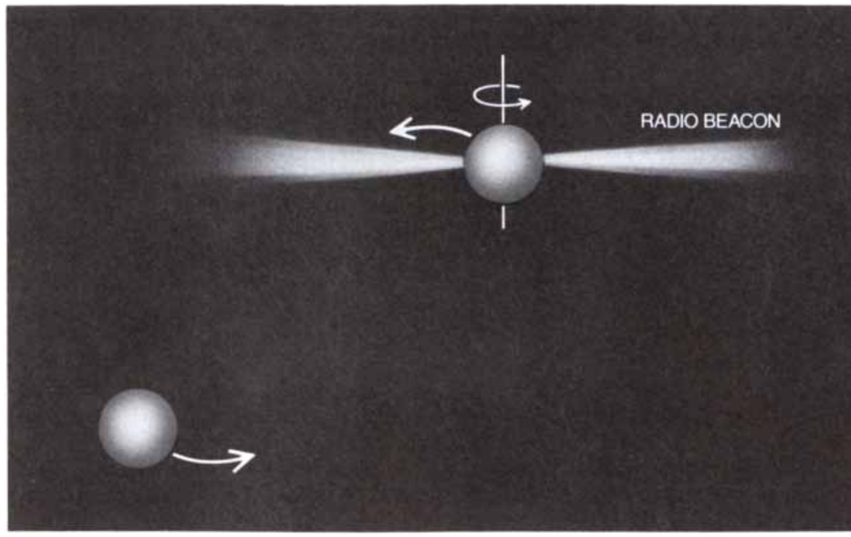
As a binary dies, its constituent stars spiral rapidly inward until they collide or are torn apart. In the case of neutron stars either event should produce a strong burst of gravitational waves. The binary pulsar will not suffer this fate for 100 million years or so, but galaxies such as the Milky Way contain enough other neutron-star binaries for several pairs to coalesce every 1,000 years. A millennium is a long time to wait, however. The LIGO is being designed so that it will be sensitive enough to detect signals from the Virgo Cluster and other nearby clusters of galaxies, perhaps recording several events per year that have a strain of one part in 10^{22}.

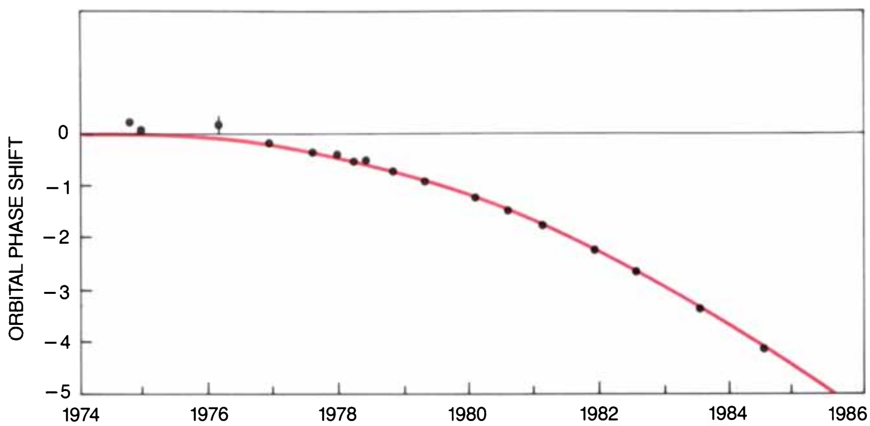
The birth of a neutron star should also be announced with a strong burst of gravitational waves. When a massive star dies in a brilliant explosion called a supernova, its core suddenly collapses into a neutron star. Although knowledge of the details of the collapse is still far from complete, it is possible that as much as .1 percent of the mass of the neutron star is converted into gravitational waves. A supernova in our galaxy, an event that occurs about once every 30 years, might create a strain at the earth of about one part in 10^{18}. The recent supernova in the Large Magellanic Cloud, a satellite galaxy of the Milky Way, should have produced gravitational wave strains of about one part in 10^{19}, just below the threshold of the best existing detectors. Supernovas in the Virgo Cluster, which appear several times a year, could cause strains of one part in 10^{21}.

The detection of gravitational waves from a supernova would have far-reaching consequences. Investigators will be interested in measuring how much time elapses between the detection of gravitational waves from the collapsing core of a supernova and the arrival of light waves from the exploding outer layers of the star. If the gravitational waves and the light waves are detected simultaneously, it would directly confirm the prediction of relativity that gravitational waves travel at the speed of light. If a supernova in the Virgo Cluster were seen in optical telescopes within a day of the detection of the gravitational pulse, then the speeds of the two signals would match to within one part in 10 billion.

Gravitational waves generated by a supernova would also provide a new window of observation. Electromagnetic radiation produced deep in a collapsing core is trapped by the outer layers of the star, hiding the most violent action from view. The detection of a burst of neutrinos from the recent supernova in the Large Magellanic Cloud gave strong evidence for the core-collapse model, but even the neutrinos were scattered thousands of times on their way out of the star. Gravitational waves, whose interaction with matter is so weak that they can escape unattenuated through the stellar atmosphere, could reveal fine details of the collapsing core. Gravitational radiation could also make it possible to see objects obscured by interstellar dust, such as the center of the galaxy.

A middle-aged neutron star could also be a source of gravitational waves, if the star rotates about an axis lacking perfect symmetry. If the hard





FIRST EVIDENCE OF GRAVITATIONAL WAVES comes from a binary pulsar discovered in 1974 by Joseph H. Taylor, Jr., and Russell A. Hulse, then at the University of Massachusetts at Amherst. A binary pulsar consists of two stars orbiting their common center of mass (*top*), where one of the stars is a pulsar: a radio source that appears to flash on and off as it rotates about its axis. The binary pulsar has an orbital period (eight hours) that is slowly decreasing (*bottom*). The decrease, which is measured by carefully timing the arrival of the pulsar radio signals, implies that the paired stars are gradually losing energy and spiraling together. The observed energy loss (*data points*) agrees accurately with the loss predicted for the system through gravitational radiation (*curve*); gravitational waves would carry away just the right amount of energy and angular momentum.

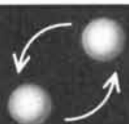
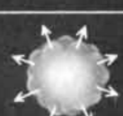
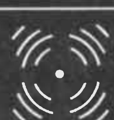
crust of the star froze in a shape flattened by rotation, for instance, an encounter with another star could shift the rotation axis. The resulting wobble should be large enough to produce strong gravitational waves.

The formation of a black hole or the collision of two black holes should also generate strong gravitational radiation. Black holes are incredibly dense objects whose existence is predicted by general relativity. A black hole is thought to form when a dead star too massive to sustain itself collapses under its own weight, leaving no trace except its electric charge, its spin and its intense gravity—so powerful that not even light can escape its pull. If a large fraction of the mass in the universe is hidden in the form of black holes, collisions are likely to be frequent enough for any detector seeing neutron-star events to observe many black holes as well. Some binary star systems that produce X rays are believed to contain black holes, although without a gravitational probe it is impossible to be sure. Gravitational wave observatories would provide such a probe, because the gravitational radiation a black hole should emit would have a characteristic signature. The detection of the signals would provide the only known unambiguous way to demonstrate that objects with these strange gravitational characteristics exist.

Another speculative source of gravitational waves is the big bang: the primordial explosion some 15 to 20 billion years ago that gave birth to the universe. Some of the most important information about the early universe comes from observations of the cosmic microwave background, the remnant of thermal radiation that filled the universe at its beginning. The detection of a background of gravitational waves would reveal new aspects of the big bang. For example, quantum fluctuations from the first 10^{-43} second of the universe might have left an imprint that could be observed today. Relict gravitational waves could also reveal the violent beginnings of galaxy formation.

The gravitational waves produced in any of the above scenarios would leave only faint and ephemeral trails as they pass by the earth. Masses would be lightly jostled, shifting positions by perhaps 10^{-21} meter (one-millionth the diameter of a proton) per meter of separation; pulsed waves, perhaps from a supernova, would last for only a few milliseconds. How could such faint signals be detected?

The first detectors for gravitational waves were pioneered in the early 1960's by Joseph Weber of the University of Maryland at College Park. The instruments consist of single massive metal bars. Each bar is a solid cylinder, typically made of aluminum. The bars are constructed to vibrate in response to gravitational waves. Currently the best bar detectors are sensitive to strains as small as one part in

SOURCE	SIGNAL TYPE	FREQUENCY	STRENGTH
 STELLAR BINARY	PERIODIC	1 MEGAHERTZ OR LOWER	10^{-21}
 NEUTRON-STAR BINARY	QUASIPERIODIC	SWEEPS TO 1 KILOHERTZ	10^{-22}
 ACCRETING NEUTRON STAR	PERIODIC	200 TO 800 HERTZ	3×10^{-27}
 TYPE II SUPERNOVA	IMPULSIVE	1 KILOHERTZ	10^{-21}
 VIBRATING BLACK HOLE	DAMPED SINUSOID	10 KILOHERTZ FOR ONE SOLAR MASS 10 HERTZ FOR 1,000 SOLAR MASSES	?
 GALAXY FORMATION (BY COSMIC STRINGS)	NOISE	BROAD BAND 1 CYCLE PER YEAR 300 HERTZ	 10^{-14} 10^{-24}
 BIG BANG	NOISE	?	?

GRAVITATIONAL WAVES from a number of sources could be detected. As the waves pass by the earth they would lightly jostle masses; waves traveling from a stellar binary, for instance, would shift all masses by a mere 10^{-21} meter per meter of separation.

10^{18}; no signals have been detected at this level.

It is possible to make more delicate measurements with the bars, but it will be difficult. One reason is that the sensitivity of any detector of gravitational radiation increases with its length, and the largest practical bar would be tens of meters long. In the approach that we and others are taking instruments that have a length of several kilometers can be constructed. The instruments we are building are called interferometric detectors. They consist of separate masses connected only by beams of light.

Interference is associated with all kinds of waves, including electromagnetic waves. Water waves illustrate the principle. Imagine two identical sets of water waves generated by pebbles thrown into a pond. The waves travel toward one another on the surface. The height of where the ripples cross is the sum of the heights of the individual waves. When a crest of one wave meets a crest of another, the resulting crest is bigger; when a crest meets a trough, the waves vanish.

Interference can be exploited to measure the relative distance traveled by two waves. The waves will add up if they have traveled precisely the same distance, but they will cancel if their paths have differed by half a wavelength. The effect provides the basis of interferometric gravitational wave detectors. Two identical light waves are prepared by employing a beam splitter—a partially reflecting mirror—to split a single laser beam into two perpendicular beams. The two beams are then aimed at suspended masses free to respond to gravitational strain. Mirrors mounted on the masses reflect the beams so that they are reunited at the beam splitter. The brightness of the resulting output beam, which is proportional to the square of the combined wave amplitude, indicates the difference between the distances to each of the masses. If a vertically traveling gravitational wave were to impinge on the apparatus, the distance one beam travels would be shortened and simultaneously the distance the other beam travels would be lengthened. As a consequence the brightness of the output would change. If the brightness were measured carefully, gravitational shifts much smaller than a wavelength could be discerned.

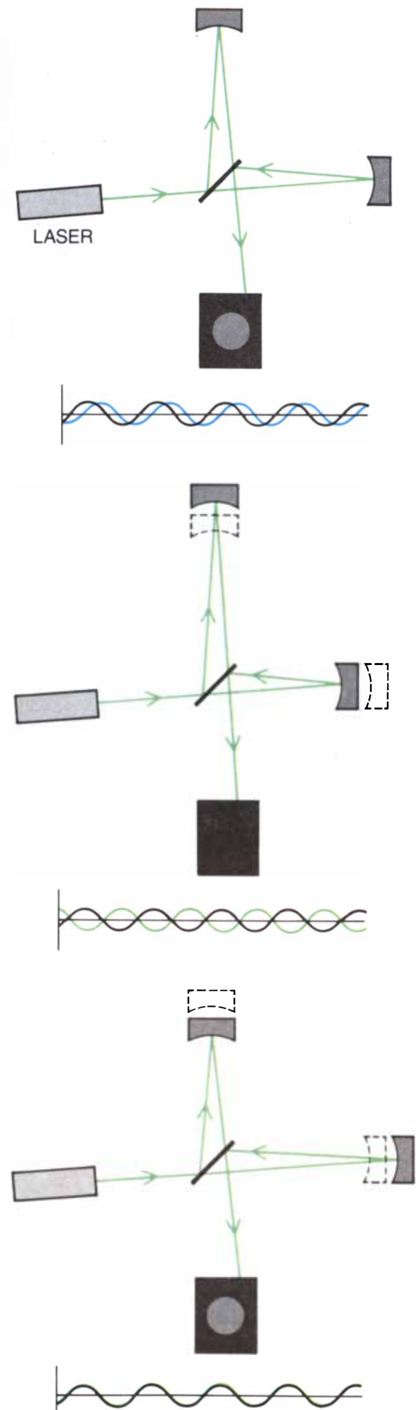
A limitation on how precisely the brightness can be measured comes from the particle nature of light. Photons (the particles of light) are emitted rapidly by the laser illuminating the interferometer. The fractional error in counting photons is equal to the inverse square root of the number detected during the measurement, so that a more powerful laser can reduce the error. The smallest one-millisecond jump measurable in a simple interferometer illuminated by a one-watt laser is 10^{-14} meter. Interferometers can measure displacements much smaller than an atom, even much smaller than the nucleus of an atom, because the light beams are reflected by trillions of atoms on the mirror surface.

The potential of increasing an interferometer's length is a key to increasing its sensitivity. Current prototype interferometers range in length from 1.5 meters (at M.I.T.) to 40 meters (at Caltech). The test masses of the interferometers planned for the LIGO will be four kilometers apart, 100 times greater than the Caltech experimental prototype.

The sensitivity of all current and planned detectors is also enhanced by bouncing the light back and forth many times between the interferometer mirrors. The improvement in sensitivity increases linearly with the number of bounces, unless the light spends so long in the arms that the effects of the gravitational waves average out. Two different methods can be used to "fold" the light path [*see illustration on next page*]: an optical delay line (suggested by Weiss) and a Fabry-Perot cavity (suggested by Drever). In an optical delay line light passes through a small hole in a mirror adjacent to the beam splitter and is reflected repeatedly before it reemerges from the entrance hole. In the Fabry-Perot cavity light passes through a partially transmitting mirror and builds up in a resonant cavity capped with a totally reflecting mirror. The light subsequently leaks out through the entrance mirror.

To search for signals near a specific frequency, the light storage time can be matched to half of the gravitational wave period. The sensitivity is then increased by exchanging the light between the two arms at the search frequency. Better sensitivity can also be achieved by recycling the light that emerges from the interferometer. It is fed back into the instrument, effectively increasing the available laser power.

The effect of gravitational radiation is universal: all masses feel gravitational waves. The test masses at the heart of gravitational wave interfer-



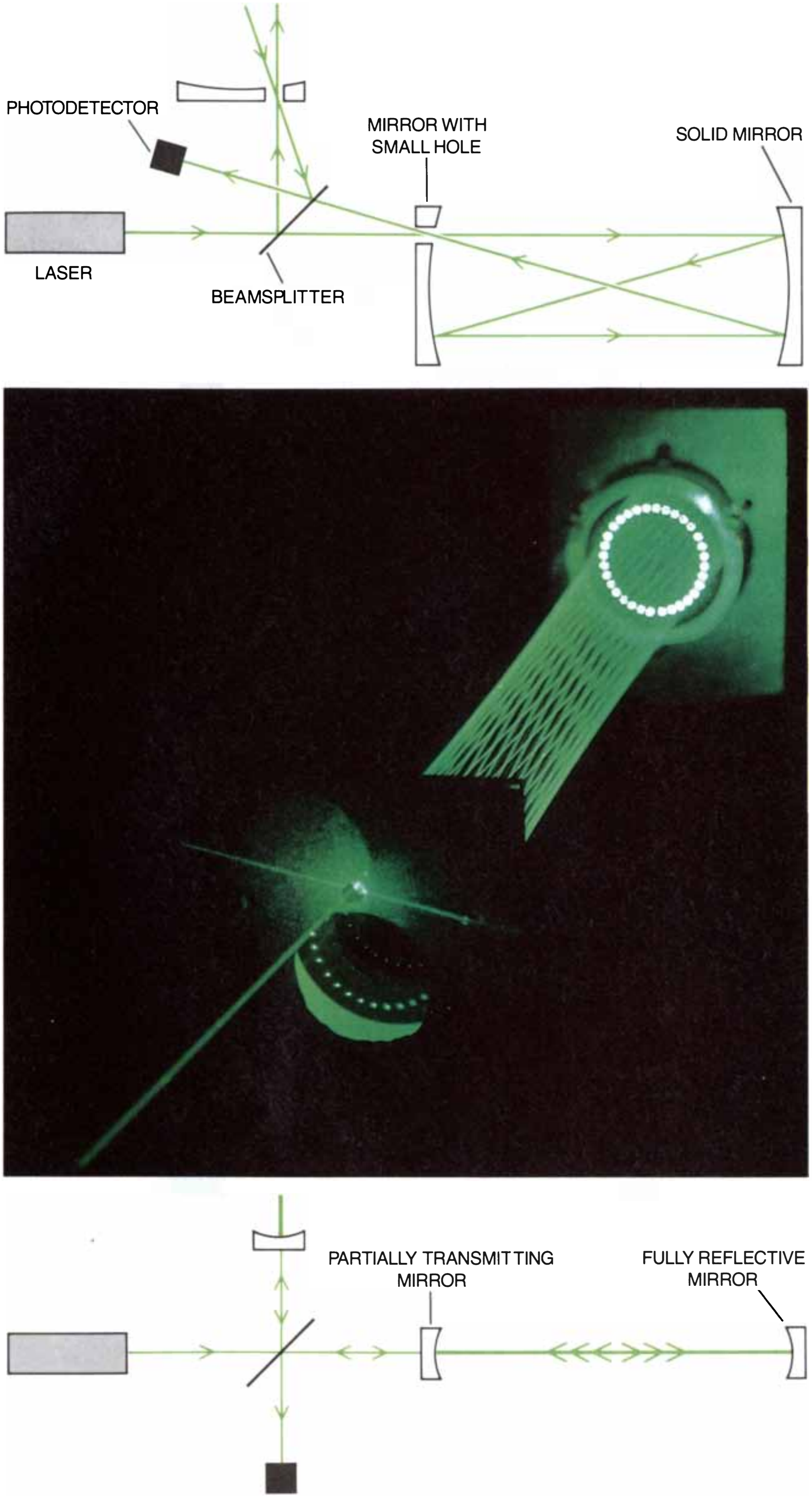
INTERFEROMETER could be employed to detect gravitational waves. A simple interferometer consists of a laser, a beam splitter, two mirrors defining two arms at right angles and a photodetector. Here a graph of the electric field in each returning light beam as a function of time is shown next to the photodetector. If the two arms differ in length by an eighth of a wavelength, the electric fields in the light beams recombine a quarter cycle out of phase with each other, and half of the light appears at the output port (*top*). If a gravitational wave shifts the path length in one sense, the light interferes more destructively at the output port: the output is less (*middle*). A shift in the opposite sense causes the light to interfere constructively, so that the output is brighter (*bottom*).

ometers are exceptional in only three ways. They are free to respond to a passing wave; they are embedded in an instrument that is sensitive to the minute displacements induced by such a wave, and they are isolated from forces that could overwhelm the signal.

The large catalogue of unwanted noise that threatens to mask the masses' response to gravitational radiation includes seismic disturbances, thermally excited vibrations of the masses and the effects of residual gas molecules striking the masses and penetrating the laser beams. Background vibrations assault terrestrial gravitational wave observatories with displacement noise many times larger than the expected signals. Large motions of the ground, as triggered by earthquakes or the occasional nearby activity of people and machinery, are infrequent and identifiable as terrestrial in origin. More troublesome is the constant inescapable background of low-amplitude seismic activity, driven by wind and waves and by the earth's internal motions.

The strength of background seismic noise drops rapidly with frequencies above 10 hertz, but even at several hundred hertz it far exceeds the required detector sensitivity. To isolate gravitational wave observatories from such unwanted effects, the test masses are suspended like pendulum bobs on thin wires. Fast motions of the suspension point are greatly attenuated before they reach the mass below. The suspension points in turn are isolated by a springy support structure, which acts like an automobile suspension system to smooth out the bumps of seismic activity. This combination of wires and springs appears adequate to block seismic noise having a frequency as low as 500 hertz; more sophisticated suspension systems based on active feedback loops may be necessary to look for gravitational waves having lower frequencies.

Another source of noise is from thermally excited vibration of the test masses. The masses ring incessantly, driven by their own internal energy. The interfering effect of thermal ringing is lessened by making the masses compact and structurally simple, so that they resonate at frequencies well above the band of expected gravitational waves. The corner area (where the vacuum tubes intersect) of most interferometers contains the bulk of the complex optics, which, if they were concentrated on one mass, would result in a structure with many overlapping resonances. In practice the central mass is usually divided into separately suspended components, each with a tolerable set of resonances.



SENSITIVITY of an interferometer to changes in the separation of its mirrors—and hence to gravitational waves—can be increased by sending light beams back and forth many times. There are two related multiple-pass configurations. In an optical delay line (*top*) light enters through a hole in a mirror, undergoes a series of reflections and leaves through the same hole. Two round-trip passes are shown here for simplicity; delay lines in gravitational wave detectors make dozens of crisscrossing passes (*middle*). In a Fabry-Perot cavity (*bottom*) light enters through a partially transmitting mirror, bounces off another mirror and returns to the input mirror. There a fraction of the light is transmitted and the rest is reflected back to repeat the cycle. The number of bounces depends on the reflectivity of the input mirror; up to 10,000 bounces have been achieved in prototypes.

Announcing the most dramatic development in home dental care since the invention of the toothbrush.

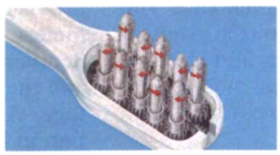
Since early man invented the first toothbrush, the technique for using it has remained just as primitive. So primitive that even today, 9 out of 10 Americans end up with some form of gum disease. The problem, historically, has been how to remove plaque. The solution now comes in the form of a technological breakthrough called the INTERPLAK® Home Plaque Removal Instrument.

INTERPLAK cleans teeth virtually plaque-free.

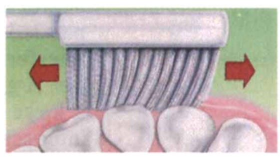
Plaque is the real villain in oral hygiene. If not removed daily, its bacterial film can lead to early gum disease and tooth decay. But clinical studies have shown that manual brushing removes only some of the plaque buildup. Those same studies, on the other hand, show INTERPLAK cleans teeth and gums virtually plaque free.

How INTERPLAK cleans circles around ordinary brushing.

With manual or even electric toothbrushes, you move the bristles up and down or back and forth. But with INTERPLAK's patented design, the brush remains still while the bristles rotate. Ten tufts of bristles literally scour off plaque and stimulate your gums. And at the precise moment they reverse direction, the tufts fully extend to clean deep between teeth and under gums. Yet because the bristles are four times softer than the softest toothbrush, INTERPLAK is no more abrasive than manual brushing with toothpaste.



Each bristle rotates 4,200 times a minute and reverses direction 46 times a second.



Hand brushing's back-and-forth and up-and-down motion often misses critical areas where plaque breeds.

The counter-rotational movement of INTERPLAK tufts clean plaque from between teeth and under gum line.

Dental professionals approve.

Dentists across the country are praising the effectiveness of INTERPLAK.

"I am recommending INTERPLAK to all my patients." –Dr. L. K. Yorn, Cedar Grove, NJ

"At last, my patients enjoy using a product we recommend." –Dr. J.W. Blackman, III, Winston-Salem, NC

"Since my patients have been using INTERPLAK, I have seen a dramatic improvement in the health of their teeth and gums." –Dr. S.G. Newhart, Orthodontist, Beverly Hills, CA

"A technical breakthrough in home dental care." –Dr. A. Kushner, Chicago, IL

Ask your own dentist about the benefits of using INTERPLAK.

The future of home dental care has never looked brighter.

Each INTERPLAK Home Plaque Removal Instrument comes with two interchangeable brush heads. You can purchase additional brush heads so every family member can benefit from cleaner teeth and gums. INTERPLAK is also cordless, recharging between uses in its own stand.

For more information on this remarkable instrument, call toll-free **1-800-537-1600, Operator 98.**

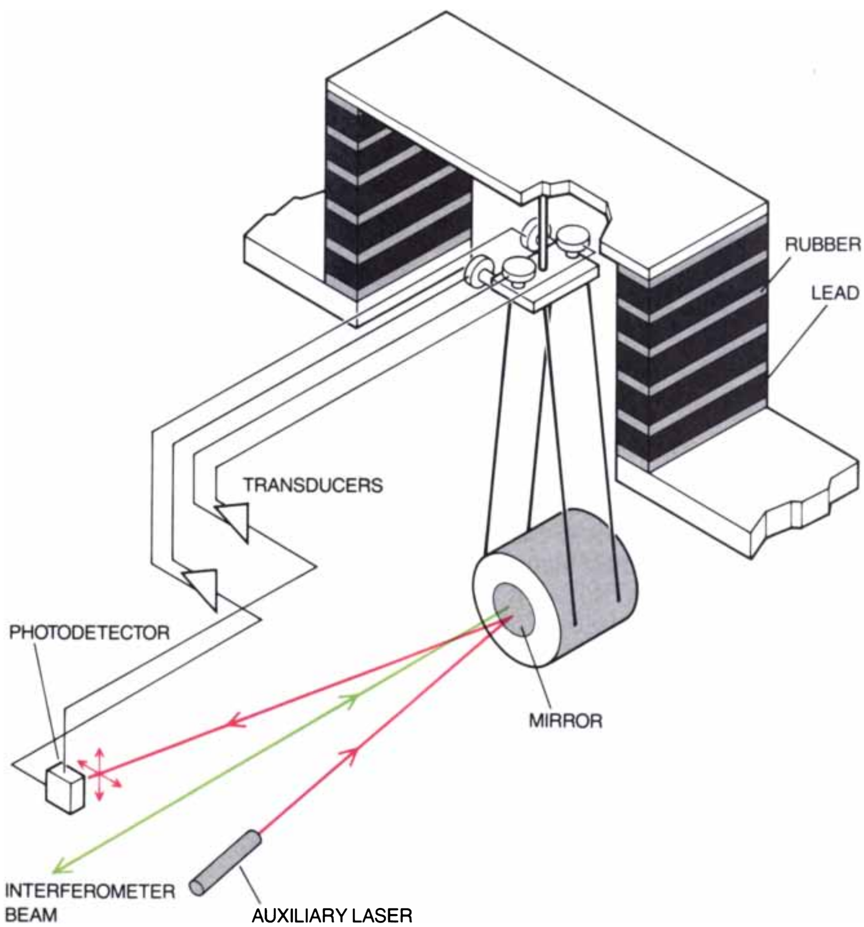
INTERPLAK®
HOME PLAQUE REMOVAL INSTRUMENT

Interplak is Acceptable as an effective cleaning device for use as part of a program for good oral hygiene to supplement the regular professional care required for good oral health.
Council on Dental Materials, Instruments and Equipment, American Dental Association

The effects of residual gas molecules striking the masses and penetrating the laser beams are minimized by the vacuum tubes. Since the speed of light fluctuates as it travels through residual gas, a nearly-perfect vacuum is required.

Smaller disturbances arise from such phenomena as changing local magnetic and gravitational fields. But few noise sources, known or unknown, are likely to simultaneously affect detectors separated by thousands of kilometers, such as the pair of detectors in the LIGO. Gravitational radiation from cosmic sources uniformly bathes the earth and will similarly and almost simultaneously affect detectors anywhere on the earth. The signature of a pulse of gravitational radiation, simultaneously registered in two or more widely separated detectors, would be difficult to forge with noise.

Current laboratory-scale interferometers lack the sensitivity to detect gravitational waves from distant galaxies; sources within the Milky Way that could be detected erupt infrequently—perhaps once every 30 years. With its four-kilometer arms, the LIGO should be sensitive enough to register bursts with strain amplitudes smaller than 10^{-21}, a level at which many astrophysical signals are expected. Moreover, almost all noise that interferes with measuring the displacement of test masses, including seismic noise and thermal motion, is independent of the length of the arms, in contrast to the strength of the registered signal, which increases with length. Of the noise sources that would increase with length, such as fluctuations in pointing the lasers and deflections and attenuations from gas molecules, most are inherently small or easy to control.



TEST MASS in an interferometric gravitational wave detector must be isolated from unwanted external forces such as seismic disturbances. That is done by suspending each mass, which can be just an interferometer mirror itself, from very thin wires. Additional isolation is provided by attaching the other ends of the wires to a structure made from alternating layers of lead and rubber. To maintain the angular orientation of the mirror, an auxiliary laser beam is reflected from the mirror's surface onto a position-sensing photodetector. The photodetector sends signals to an electromagnetic transducer that tilts and turns a smaller mass near the suspension point. The smaller mass acts like a puppeteer's handle to correct any angular deviations that might interfere with measurements.

A single gravitational wave detector has only weak directionality. This means that interferometers in different parts of the world can simultaneously register signals from a large area of the sky. Although two detectors are adequate to distinguish bursts of background noise, at least three operating detectors will be necessary to measure precisely the direction from which the burst signals come. The information about direction would result from comparing the arrival times of the burst in the various detectors. Two European projects to build large interferometers have roughly the same schedule as the U.S. plans. One, a detector with one-kilometer arms and provisions to expand to three kilometers, is planned for construction in Scotland by the University of Glasgow and the Rutherford Appleton Laboratory. Another, planned for construction in Bavaria by the Max Planck institutes in West Germany, will have three-kilometer arms meeting at 60 degrees; the observatory could be expanded to accommodate three nested interferometers in an equilateral triangle. A group in France is also planning a large interferometer, although on a later schedule.

The LIGO will have vacuum pipes that are large enough (1.2 meters in diameter) to accommodate several interferometers, each interferometer with separately suspended test masses. Interferometers running half the length will respond just as the full-length interferometers do to most noise influences but only half as much to legitimate signals. The approach will distinguish between the signals and spurious disturbances.

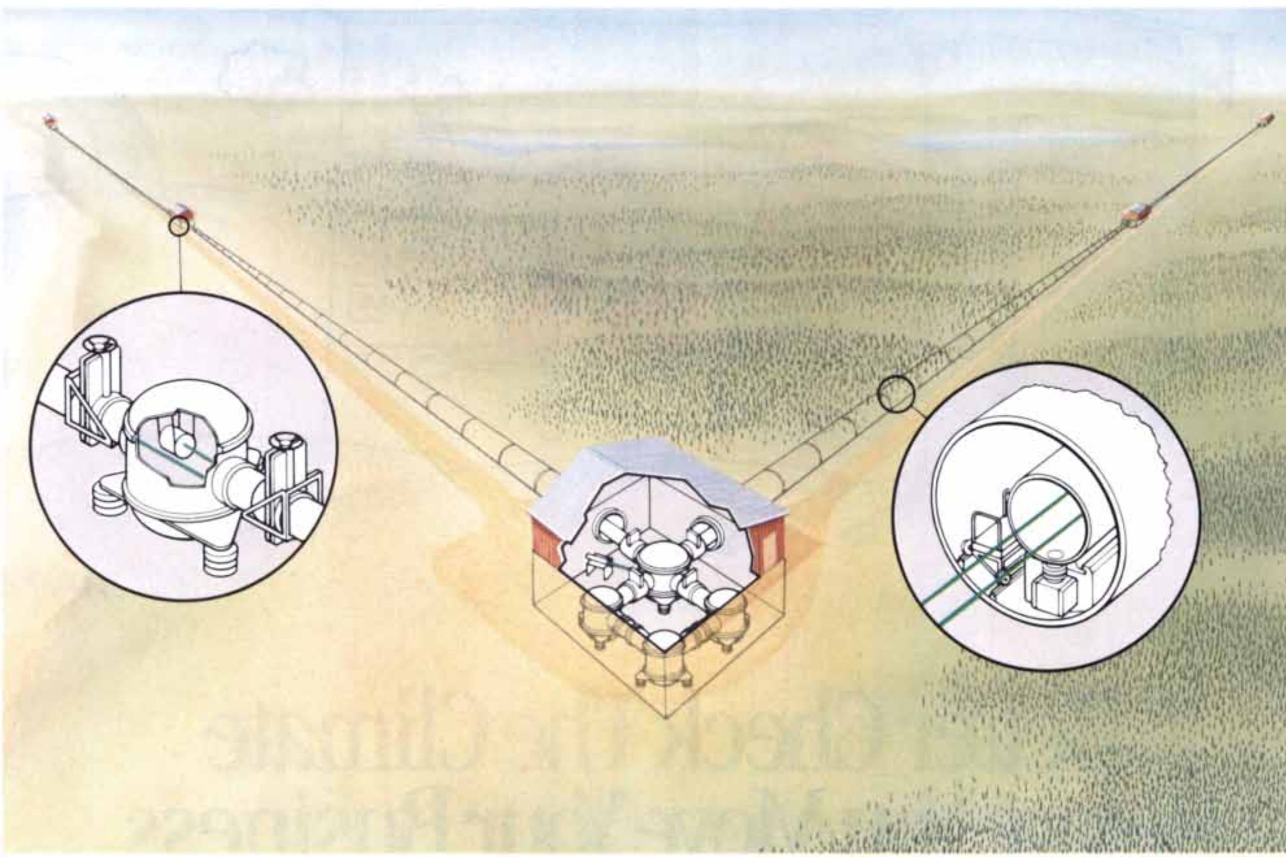
Since much of the noise that interferes with measurements comes from the terrestrial environment, it would make sense to set up gravitational wave observatories in space. The idea is being actively pursued. One type of experiment involves precisely measuring the frequency of a microwave signal that has been returned to the earth from a deep-space probe. A gravitational pulse would shift the frequency of the wave. The method is sensitive to gravitational waves in the frequency range of one per minute to one per hour, which is well below the frequency range of any terrestrial antenna. The sensitivity so far achieved has been insufficient to detect any waves.

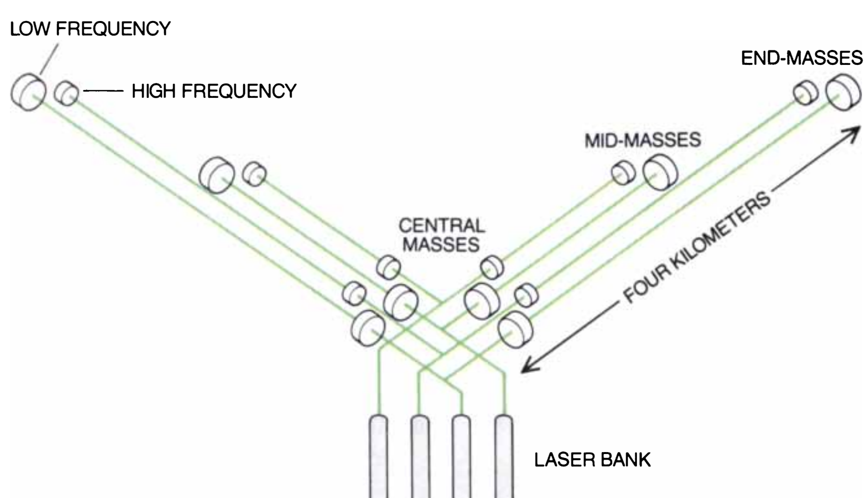
A more elaborate space experiment now under consideration would employ lasers to measure the separation of test masses placed in orbit about the sun. The distance between the masses could be as much as a million kilometers. Such a system would be well suit-

ed for studying the gravitational waves from classical binary stars, which revolve slowly. Although a much shorter space interferometer might be useful in the kilohertz band, it would probably not perform any better than terrestrial detectors at those frequencies and would be considerably more expensive to build.

Each advance in the technology of astronomy, including radio, X-ray, infrared and advanced optical telescopes, has brought a new dimension to the sketchy image of the cosmos. The 19th-century picture of the universe was one of a peaceful and timeless collection of stars. Now it is known that the universe was born in a titanic explosion and is expanding at enormous speed. The universe is itself filled with island universes, which are in turn the seat of such violent events as gas jets roaring across thousands of light-years, old stars being torn apart by their neighbors and new stars igniting inside vast dark clouds.

We expect that the LIGO and the other gravitational wave observatories will provide the first clear view of processes that can now be only glimpsed with electromagnetic radiation. The observatories could very well detect the collapse of cores of supernovas, the relativistic coalescence of neutron stars and perhaps even the formation of black holes. Most important, our understanding of the universe will once again be enriched by the discovery of completely new phenomena—visible only by their gravitational radiation.

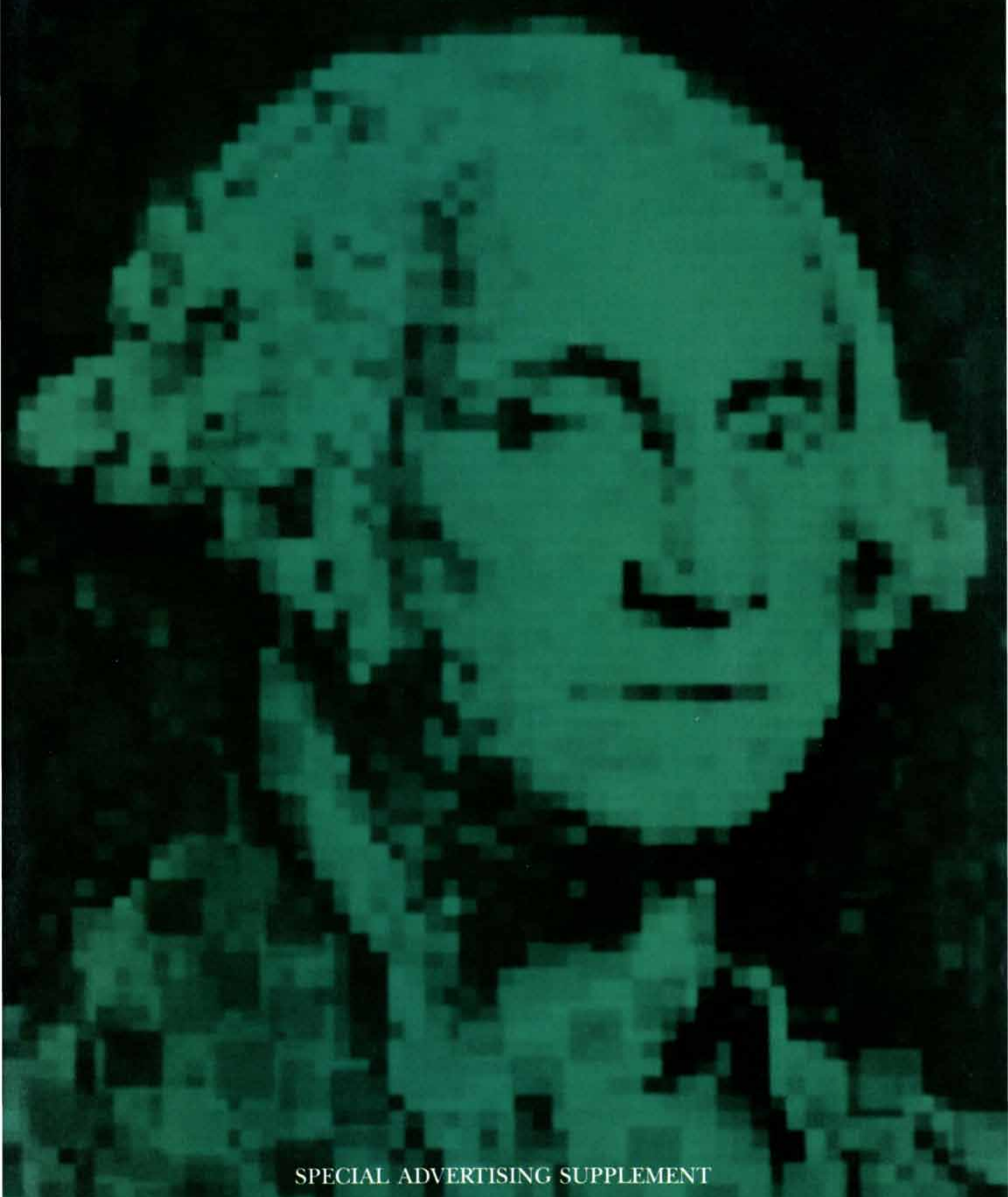




LARGE INTERFEROMETER scheduled for the 1990's is actually one of a pair of interferometers known collectively as the laser interferometer gravitational wave observatory, or LIGO. One of the interferometers is to be built in southern California and the other in Maine; the two instruments will be linked electronically to operate as a single observatory. Shown here is an artist's representation of one of the detectors. Each instrument will have arms four kilometers long. One possible arrangement for the LIGO has four interferometers, each powered by a separate laser, sharing one set of pipes (*right*). Such side-by-side operation of interferometers, each optimized for a different astrophysical signal, would increase the LIGO's versatility. Each optical system is duplicated in a half-length (two-kilometer) configuration so that spurious signals can be rejected.

THE SCIENCE OF MAKING MONEY

Science and Technology Revolutionize Wall Street



SPECIAL ADVERTISING SUPPLEMENT

SCIENCE AND TECHNOLOGY REVOLUTIONIZE WALL STREET

◆

by Dexter Senft
The First Boston Corporation

Wall Street in 1987 is a curious beast, blending old and new, art and science. Investors, dealers, brokers, analysts and speculators go about the game of making money with the same basic strategies used for centuries: buy low and sell high; profit is good, risk is bad. But the equipment, the tactics, and even the rules of the moneymaking game have changed radically. The players show up with a dazzling array of technology, from super-computers to portable micro-processors, vast data bases and high-speed communications networks. Many sport fancy educations as well. The usual plethora of MBAs is peppered with PhDs in physics, applied mathematics, computer science and the like. And just when the players begin to get comfortable with the game, a breakthrough in technology will make a tactic obsolete or a new law or regulation will render it moot.

To understand the new role of technology on Wall Street, begin by looking at the three assets of a securities dealer or investment bank: people, capital, and information. All are highly leveraged.

A typical "major" investment bank in the United States has over $1 billion of capital and about 5,000 employees. The capital is used to acquire some $30-$40 billion of assets, the difference being borrowed. The people are used to buy and sell these assets, continuously depleting and replenishing them at a "turnover" rate of about once per day. This means that total annual purchases and sales work out to about $10 trillion, or about 2.5 times the United States GNP. With skill, luck, and good information, each sale will take place at a tiny fraction of a percent higher in price than the corresponding purchase. Needless to say, the investment banker wants the best possible information, as quickly and easily as possible. To get it, the company will probably spend upwards of $100 million per year on its technology.

With its tremendous volume and thin margin, the typical investment bank knows all to well the meaning of risk. In fact, more of its technology is focused on the management of risk than on the creation of reward. Buzzwords such as "securitization," "dynamic asset allocation," "negative convexity" and "program trading" all have foundations in risk management. The hardware, software, and brains ("wetware") brought to bear on these and other problems are nothing short of a technological revolution.

New technologies spur the creation of new money management techniques and new types of securities. Investor acceptance of these follows, as do new associated problems or risks. These risks, in turn, create demand for new risk management tools, strategies, systems, and quantitative experts. The result is a chain reaction that has produced the "explosion" of new securities so widely publicized in the last two to three years. The average fixed income securities salesman, for example, has three times as many products to sell today as he or she did five years ago.

Technology at Work

Fifteen years ago, the average securities dealer had a mainframe computer that managed its books and records. Period. The computer billed trades, maintained the securities inventory, and generated the profit and loss statements. It was a behemoth—slow, expensive, large, and very unfriendly. Most input and all programming was in the form of punched cards, the "holes" from which were saved for the next ticker tape parade. Salesmen figured commissions from a chart, bond traders looked up yields in books, and finance officers projected income statements on spreadsheets—the wide, green paper kind.

Today, that much computing power costs about $5,000. The CPU is a circuit board that fits in a briefcase, and each high-density floppy disk that it reads and writes is the equivalent of 15,000 punched cards, except that the floppies can be erased and reused. Meanwhile, the multi-million dollar mainframe is tackling in minutes complex analyses that would have taken months, working off data bases that 15 years earlier might not have fit in the entire office building. More than any other factor, cheaper and faster computers have revolutionized the securities business.

Other technologies have played a role, too, although they are often paired with cheap microprocessors. On a walk through a modern trading floor, one could easily spot technologies that were non-existent only five years ago, let alone 15. The most obvious of these is the video and digital switching network that drives the entire floor. Each trader and salesman views several video screens, all controlled by a single keyboard. From the keyboard the user can summon any of hundreds or thousands of information "pages" from any of the

---

Wall Street is undergoing a technological revolution. Just as nineteenth century factories turned to water-power to boost production, investment banks and stockbrokers are turning to computers and high-speed telecommunications links to speed the flow of money and securities. But Wall Street's information revolution goes a step further. Number-crunching computers now allow Wall Street firms to create entirely new financial products: new risks and new ways to hedge against them.

firm's computer systems, or a personal computer.

Increasing amounts of this information are in digital format—which means that it can be processed just as any other computer data might be. (The rest is in video format, which can only be displayed.) Microprocessors can be easily programmed to automatically monitor all incoming news and price quotations and to select those deemed to be interesting to the user. Traders can watch for volume limits or spread targets; salesmen can monitor situations their accounts care about. Since most news services are in digital format, stories can be saved and later scanned for key words or phrases.

The Information Stream

Digital information is nothing new to the stock market. In fact, the ticker of an exchange is a digital feed that has been available for decades. These ticker streams are now broadcast using sideband frequencies of various FM radio stations; signals can be monitored with special receivers that can be carried in a briefcase or installed in a personal computer. The market that will be revolutionized by digital information is the bond market. Most bonds trade "over the counter"; that is, the trades do not appear on any exchange or ticker. Brokers and dealers have long maintained information pages for bonds in video format. As these become digitized, they provide the first real-time price stream for the fixed income market.

Another important use of digital price information is in fine tuning risk management. Many firms now monitor their trading positions minute by minute. Combining split-second bookkeeping with real-time, digital prices allows positions to be continuously monitored and hedged. Better position and hedging information makes securities dealers more comfortable with larger trading positions and inventory, and ultimately adds liquidity to the market.

Beyond the trading floor, in the research and finance areas, the most obvious change of the past decade is the profusion of personal computers. Electronic spreadsheet programs revolutionized the corporate finance field by automating "gruntwork," thereby improving productivity by several orders of magnitude. The main drawback to PCs is their inability to store large quantities of data. Within the investment bank, this can be mitigated by uploading and downloading data to and from various mainframes over high-speed communications networks. Of course, each generation of PCs is endowed with larger and faster hard disks, so it is unwise to bet on storage capacity remaining a handicap for very long.

Mainframe computers, meanwhile, are taking on tasks too complex for smaller machines. Certain financial problems are so complex that without a super-computer they would not be worth doing. Examples include structuring large fixed income portfolios to satisfy various cash flow constraints or certain complex option valuations. Today's super-computers are capable of tackling such problems in a reasonable time frame—for example, 15 minutes or so. Any longer, and the risk that the market would change (thereby invalidating the solution) would be too great. The advent of floating point vector processors and various parallel processor technologies offer the possibility of solving even more complex problems (which are abundant) or handling the existing problems more quickly and cheaply. ◆

GLOBAL BANKING TECHNOLOGY

◆

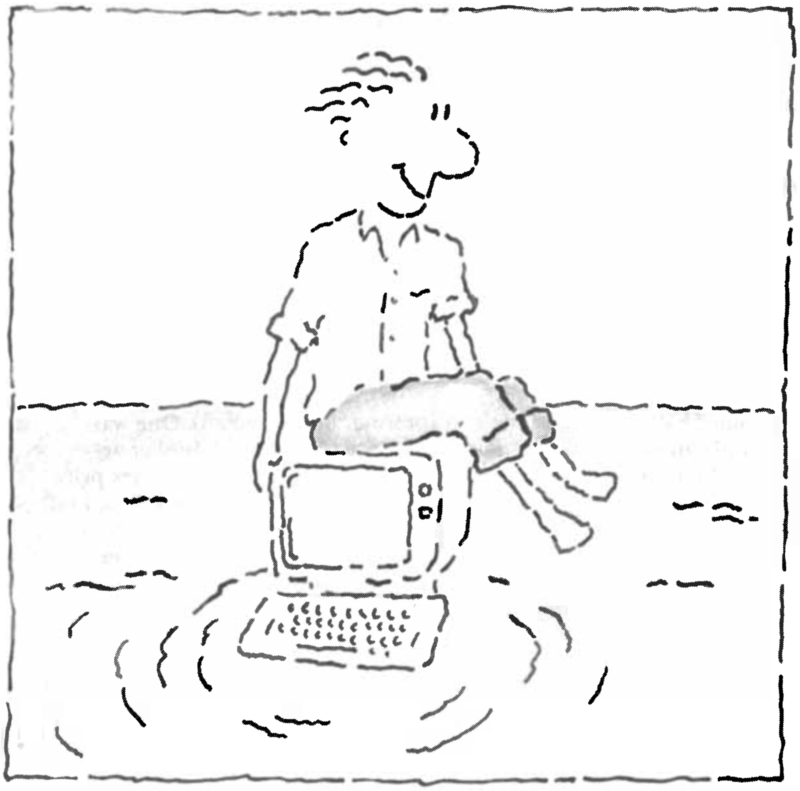
by Paul Glaser
Citicorp

Having just read a detailed account of the Japanese government's emergency trade-deficit-reduction plan, a company treasurer of the mid-1990s decides to adjust his investment portfolio to take account of the plummeting yen. Turning to his computer terminal, he revises the instructions previously given his automated investment program about yen exposure. Within a few seconds, the computer flashes up a list of securities to be traded or hedged, and the optimal places to do the deals (in this case, buying in London and selling in Tokyo). An additional keystroke tells the machine to execute the trades, and then to advise on key indicators to watch, and trading strategies to think about for the future.

Science fiction? Maybe not. Companies are already laying the foundations of such systems. Their components include telecommunications (both long-distance links and the complex switching needed to feed traders' workstations), computers, and the software needed to make these circuits do useful work. Although not as "intelligent" as the systems envisioned for the future, such electronic systems are already crucial for the smooth workings of the securities business. In the course of a normal deal a trader may tap into several. Each will have cost its developer up to $30 million.

Citicorp began creating its global communications network in the nineteenth century—before the telegraph, when packet ships carried messages now sent by packet switching. The reason for going abroad was to assist American firms operating overseas and to finance American trade. Today, Citicorp still finances global trade. But the technology has changed beyond all recognition.

GET YOUR FEET WET.



Sample Dow Jones News/Retrieval® — the nation's premier online resource for investors—without getting in over your head.

For only $29.95, the News/Retrieval Membership Kit lets you plumb the depth of this valuable business and financial information service. Five free hours* of online time will get you started.

There's a User's Guide to show you the way, step by step. And a year's subscription to *Dowline* (the magazine of News/Retrieval) to give you background on and suggested applications for our offerings.

Stay on top of stock performance, past and present. Keep up with breaking news or call up past news from *The Wall Street Journal* (online exclusively with News/Retrieval) back to January 1, 1984. Run through earnings and economic forecasts, and reports from major brokerage houses.

You can even plan a trip, catch a sports score, and tap a pool of general knowledge—using almost any computer and modem, terminal or communicating word processor.

The Dow Jones News/Retrieval Membership Kit. It can help you embark on a wave of investment success. . .and not feel soaked.

To order or for more information, call **1-800-221-7700, Ext. 454F.** (Alaska, Hawaii and foreign, call 1-215-789-7008, Ext. 454F.)

Or use the coupon.

YES! I WANT TO GET MY FEET WET!

☐ **Personal offer-$29.95 Each** includes:
- 5 Free Hours* • Single Password
- User's Guide • 1 Year of *Dowline*
- $12 Annual service fee waived for 1 year

☐ **Corporate offer-$49.95 Each** for multiple users, includes: • 8 Free Hours* • Multiple Passwords • User's Guide • 1 Year of *Dowline* • $12 Annual service fee waived for 1 year

☐ **Send more information**

$____ENCLOSED (Check/Money Order) ☐Bill me later Charge to: ☐AmExp ☐ MC ☐ VISA

(If you pay by credit card, all subsequent usage charges will be billed to your credit card.)

Card No.______________________ Exp. Date__________

Signature ______________________

Name ______________________

Company ______________________

Address ______________________

City/State/Zip ______________________

Daytime Phone Number ______________________

Mail to: DJN/R Membership Kit, P.O. Box 300, Princeton, NJ 08543-0300
ATTN: John McGovern 2OSA0607E

*Per account. Limited to new subscribers only. Free time (5 hours per Personal offer; 8 hours per Corporate offer) must be used within 30 days after receipt of password(s). Certain databases have fees over and above usage charges which are excluded from free time offer.

Dow Jones News/Retrieval®
Fuel for your ideas.

Global Telecommunications

Telecommunications is the backbone of today's global finance. Citicorp's Global Telecommunication Network (GTN) is a private data communications network developed in the early 1980s to support global electronic banking services. The network serves 75 of the 94 countries in which Citicorp has offices, with access points or nodes in 149 cities connecting terminals in customers' offices, and terminals and computers in Citicorp branches.

The network provides both switched connections and nonswitched point-to-point channels. Switched connections are similar to those provided by a public telecoms network. A computer "dials" the number of another machine, and is connected to it. All customer electronic banking services are delivered via switched connections. Switched connections use terminal-to-computer (asynchronous) communication with the customer's terminal at 300 bits per second (bps), 1,200 bps, or 2,400 bps. The network automatically handles the details of matching the speed of the customer's terminal to that of the computer at the far end.

To get the most out of expensive long-distance lines, the system can pack several calls onto a single long-distance link. Network nodes capable of packing calls onto a long-distance line, and unpacking from one are available in New York, London, and Hong Kong. They could be built in other locations as required.

Nonswitched point-to-point channels within Citicorp typically link computer systems to one another. Such hardwired connections are used where the data traffic between two systems is more or less continuous, making it inefficient to keep setting up and tearing down data calls. The GTN can provide asynchronous or synchronous point-to-point channels at any speed from 300bps to 19,200bps.

The complete 75-country network consists of a three-node "backbone", with centers in New York, London, and Hong Kong, and three regional subnetworks radiating from these nodes. The backbone nodes are interlinked by three high-capacity leased circuits, forming a triangle. A single circuit failure can be bypassed automatically by routing traffic the other way around the triangle. Most of the leased circuits from the country-level nodes to the backbone nodes carry data at 9,600 bps and are backed up by the public telephone network. The total number of concurrent calls carried by the network at the busiest times is about 1,400, with about 800,000 calls being completed per month. Of these, roughly 50% represent Citicorp internal traffic and 50% traffic between Citicorp and its customers.

Perhaps the greatest achievement of the Citicorp network, however, is that few of those using it actually notice that it is there. The trader looking for, say, bond price quotations need not worry whether that information is stored on his own workstation or on a computer halfway around the world. The applications software running on his workstations will automatically make the proper connections.

Sophisticated Software

Thanks in part to the widespread access to information which the network provides, application software is becoming more powerful. Citicorp's Global Trader, for example, unifies many of the company's worldwide securities dealing operations. It is composed of four subsystems.

The first is a Treasury dealer support subsystem, designed for foreign exchange and money market traders. Traders can quickly and easily input how much they have traded with whom by pointing with an electronic pen to one of a menu of items on a digitized pad. The deal appears on-screen for confirmation, and once authorized, the trader's updated position and profit/loss figures are displayed. After the trade has been made, back-office bookkeeping computers are automatically notified so that they can prepare the appropriate paperwork. To help choose between offered deals a decision-support feature helps the trader calculate prices and yields of various instruments and compare results. Traders can communicate with each other via an electronic notice board, determine foreign exchange and money market limits and, obviously, keep track of their accounts.

Second comes a subsystem for gilt and equity market-making. Although American banking regulations forbid Citicorp from setting up shop as a securities dealer in New York, it can, and does, deal in equities and gilt, (i.e., British government bonds) in London. This system gives it the ability to perform trades swiftly and accurately, with a minimum of paperwork. Traders input and confirm deals similarly to the Treasury system. Once the deal is finalized, the screen gives the trader a real-time view of positions, portfolios, commissions, etc. It also allows traders to perform analysis—i.e., comparing a customer's current position with predefined risk limits. And the system is linked to back-office bookkeeping systems.

Last, but far from least, are two information subsystems. One was developed by a British broker acquired by Citicorp. It integrates price and rate services, economic data, and company information services into one broad decision-support system. This subsystem combines and integrates the data onto one screen and one easy-to-read format, which is menu-driven. It provides access to forecasts and historical data on 50,000 world stock prices, and 10,000 economic and 8,000 monetary statistics. The subsystem can output both text and graphics.

The second information service was developed independently of the first, largely within Citicorp— hence the partial overlap of the two information subsystems. This one carries real-time foreign exchange rates from Citibank and eight other money center banks in New York and Europe.

The next step in the development of Global Trader are analytic systems which can help a trader decide between competing deals. Eventually, such analytic systems could be made smart enough to suggest trading opportunities on their own initiative, and they might organize trades to fit in with broad corporate rules for risk and trading strategy. That should provide most of the features mentioned in the "science fiction" with which we began this article. But by then the system users will no doubt be demanding more. ◆

Electronic circuits and fiber optic cable determine the geography of the new Wall Street. Telecommunications links increasingly decide who can trade what with whom. Computer-power often decides who will get the better of the deal. Financiers often speak of the wonders which tomorrow's technology will bring.

THE BROKER'S AUTOMATED WORKSTATION

◆

by Bernard A. Weinstein
E.F. Hutton & Company Inc.

The role of the stockbroker is changing. Some brokerage firms have recently started calling their stockbrokers "Financial Consultants". This is not the first time that the stockbroker's title has changed. In the early 1960s, a stockbroker was commonly called a "Customer's Man". In the late 1960s and early 1970s, "Registered Representative" gained popularity, later being replaced in the late 1970s by "Account Executive". Now, as we enter the end of the 1980s, "Financial Consultant" has come into vogue. (In the 1990s, who knows? Perhaps "Financial Wizard".)

How is this significant? What a securities firm elects to call its brokers is a significant clue to how it is seeking to position them competitively. And each new role the broker has taken on has required a new information system.

The Customer's Man, in fact, was essentially the upstairs guy who took the client's order and phoned it to the floor broker at the stock exchange who executed it. The only information service tools available to the Customer's Man were the telephone and the ticker, the familiar paper tape strip printer that printed out stock prices as orders were executed.

Customer's Man gave way to Registered Representative. Registered Representatives did more than answer the phone and relay orders. At their disposal was a new-fangled device: the quotation terminal. By tapping away at it, they were actually able to tell their customers the current price of their stocks without going through piles of paper tape.

Our stockbroker did so well that he soon graduated from Registered Representative to Account Executive. In many firms the Account Executives' desk-top terminals were now tied into their firm's private networks, allowing them not only to send and receive messages directly, but also to access their firm's data bases.

Account Executives could tell their customers their current account balances. They could determine the most popular securities among their customers. They had at their fingertips their firm's latest research as well as outside research opinions. From the terminal, they could retrieve in seconds news on any given stock or subject. The list goes on and on.

Once again, the wheel turns and a number of firms have changed what they call their brokers—this time Financial Consultant. The broker's information support system is changing too. The idea is to allow the broker to concentrate on what *he* does best, manage his relationship with the client, by doing what *it* does best, manage information. Thus, the broker's workstation, which has already evolved from a simple quotation device to a combined market data communications and inquiry terminal, is evolving into a total integrated information center.

The workstation supports market data, communications, information management, office automation, personal computing and an array of customized applications.

At E.F. Hutton for the past ten years, we have been painstakingly building an information system architecture to support such a workstation. This architecture includes multiple levels of distributed processing systems and data bases connected by a terrestrial and satellite network. The workstation itself, which is called AWE (Advanced Workstation for the Executive), will begin to be rolled out to our 400 branch offices this year.

Until recently, a workstation that could function as a totally integrated information center for the broker seemed to be one of Wall Street's impossible dreams, not just for technical reasons, but because of cost. Now, because of the tremendous price/performance improvement in microprocessor technology the cost of workstations has dropped so much that what was once a dream is attainable.

AWE has elements of an expert system. It does not tell the broker to buy or sell a particular stock because, for example, the dollar is weak. But it does advise him whenever there is news, a research department opinion change, heavy insider buying or selling, a technical buy or sell signal, etc., and it suggests follow-up action. For example, if there is news on XYZ, it will alert him to it and then ask: Do you wish to see your clients who hold XYZ? Do you wish to see the latest research comments on XYZ?

In short, the system will act as an expert assistant who will alert the broker to developments and opportunities that he might otherwise miss and guide him through possible analytical steps. Instead of passively displaying information on demand, the workstation will become more and more an active partner. ◆

Wall Street's rocket scientists aren't the only ones with access to state-of-the-art technology. The quantitative revolution is transforming your stockbroker's desk as well. Not long ago, the only information tools available to the broker were the telephone and the ticker tape. Now, expert system technology is available in desktop systems that can alert brokers to opportunities for their clients that they might ordinarily miss. The "Customer's Man", as stockbrokers used to be known years ago, has come a long way.

ANALYTICAL MEASURES IN TODAY'S BOND MARKET

◆

by Martin L. Leibowitz
Salomon Brothers Inc

For centuries the yardstick for measuring the worth of a bond has been its yield. This fundamental investment measure is not only critical to the smooth functioning of the bond market, but also it underlies many facets of investment analysis relevant for other kinds of assets, such as equity and real estate. A thorough understanding of the mathematics of bonds and their yields takes one a long way towards understanding the mathematics of all investments.

A bond provides a cashflow that is specified in advance. The typical U.S. bond entitles the investor to a cash payment every six months and the repayment of the principal at the maturity of the bond. The standard principal amount is $1,000 per bond. A bond with an 8% coupon represents a promised payment of $80 a year (8% x $1,000), which is paid in two semiannual installments of $40 each. Although the bond is usually issued at a price near par, or $1,000, its price in the secondary market after issuance is determined by the interest rate then prevailing *(see Figure 1, page M14)*.

Yield-to-Maturity

Although price is closely related to the ultimate amount transacted, the so-called yield-to-maturity is often the primary variable used for making investment decisions about bonds. The value of the yield-to-maturity corresponds one-for-one with the bond's price on a given day. Given some discount rate, the stream of future cash flows from a bond can be equated with a dollar amount in the present via a geometric discounting process. The dollar amount is called the present value of the cash flow. For a given discount rate, the present value can be greater or smaller than the actual dollar price at which the bond is traded. The discount rate that equates the present value of the future cash flows with the dollar price of the bond is called the bond's yield-to-maturity *(see Figure 2, page M14)*.

In the marketplace, the yield-to-maturity serves as the basic standard for the comparison of bonds. All else being equal, the higher the yield-to-maturity of a bond, the more it is to be preferred. Typically, however, the investor pays a price for higher yields in the form of longer maturity, lower credit quality, reduced marketability and the like. These investment distinctions are reflected as differences in yield, or yield spreads, between different sectors of the bond market. For example, the yield spread between a bond yielding 8.00% and one yielding 9.25% is 1.25%, or 125 basis points. (A basis point is one-hundredth of one percent.) The concepts of yield and yield spread form the foundation for the determination of value within the fixed-income market.

It should be quickly noted that yield-to-maturity is not without its flaws as a benchmark measure of value. Most participants in today's institutional bond markets are well aware of its shortcomings. Strictly speaking, if one evaluates two bonds according to their yields-to-maturity, one makes what may be an unjustified assumption about the reinvestment rate obtainable for future coupon payments. Furthermore, if one assumes a constant yield-to-maturity over all future interest periods, one ignores the possible variation of interest rates over time. The concept of yield per se fails to account for the effects of taxation. Finally, the yield-to-maturity does not reflect the contingencies associated with any rights the borrower may have to pay off the loan before maturity. The exercise of such rights can lead to cash flows radically different from the ones originally scheduled.

Treasury Yield Curve

Another fundamental construct of bond analysis is the Treasury yield curve. The U.S. Treasury frequently issues securities of many different maturities. Hence for virtually every maturity, from the shortest one-day T-bill to the longest 30-year bond, one can estimate the yield value that corresponds to a potential new Treasury issue. The Treasury yield curve *(see Figure 3, page M16)* is a plot (usually smoothed) of such yields by maturity, and it has become increasingly fundamental to the analysis of all sectors of the fixed-income market.

Analysts who study the bond market have often tried to look beyond the yield curve into what they call the term structure of interest rates. By this they mean the discount rate associated not with a standard, coupon-bearing bond with its complex cash flows, but instead with a simple bond that pays only a single cash amount at a given maturity. For example, for a six-month par bond the yield rate and the so-called "spot rate" associated with the term structure would be identical, since the only cash flow takes place six months later at maturity. For a one-year par bond the yield rate would probably be different from the one-year spot rate. Before the 1980s one had to solve a series of simultaneous

---

Automation is sending Wall Street back to basics. Using computers, firms are re-examining fundamental theories for analysing the prices of securities. Their speed can put to practical use formulas whose results were outdated by the time they emerged from clerks' adding machines in the old days. Meanwhile, Wall Street's so-called rocket scientists are working on new financial theories. Many of their most useful creations concern bonds.

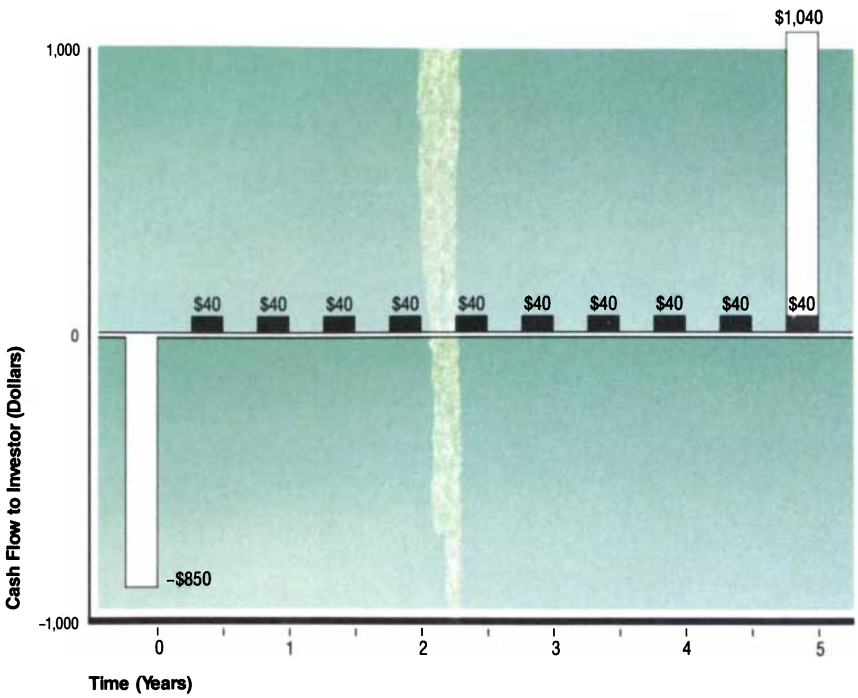
That task is accomplished with ease, too—by a 3.5-liter, 6-cylinder, fuel-injected overhead-cam engine that's capable of generating 182 horsepower.

In a perfect world, in short, everyone would drive an L6. Unfortunately, only a severely limited quantity will be imported in the coming year.

Leaving one small consolation for the vast majority of the world's driving population: They will be spared the sudden shock of ever having to step out of the BMW L6 coupe, only to re-enter a world that suffers much by comparison.

THE ULTIMATE DRIVING MACHINE.

Figure 1



Cash Flow from a five-year, 8% bond is shown schematically. The purchase price is 85, or 85% of the $1,000 face, or par, value of the bond; it is shown as a negative cash flow because from the investor's point of view it represents an outflow of funds. In return, the investor receives a coupon payment, or interest payment, every six months, which amounts to $40. Thus the two coupon payments made in one year total $80, or 8% of the $1,000 face value of the bond. At maturity the face value of the bond is returned to the investor, together with the last $40 coupon payment.

Figure 2

$$10P = \frac{10\ C/2}{(1 + y/200)} + \frac{10\ C/2}{(1 + y/200)^2} + \cdots + \frac{10\ C/2}{(1 + y/200)^{2M}} + \frac{1{,}000}{(1 + y/200)^{2M}}$$

Price of A Bond is the present value of the future cash flows the issuer promises to pay the investor. Here *P* is the price as a percentage of the $1,000 face value of the bond, *C* is the coupon rate as an annual percentage of the $1,000 face value and *M* is the number of years from the issue of the bond to maturity. In order to determine present value one must assume a discount rate *y*, which is always quoted as an annual percentage rate.

equations for par bonds with different maturities in order to determine the theoretical spot-rate curve associated with the Treasury yield curve.

In recent years, stripping the cash flow of Treasury securities has created an entirely new market, made up of instruments such as CATS (Certificates of Accrual on Treasury Securities), TIGRs (Treasury Investment Growth Receipts) and STRIPS. These new instruments are single-payment Treasury securities, and so to a large extent the spot-rate curve has become a visible and regular part of the marketplace.

A related concept is the forward-rate structure. The forward rate is the interest rate that would have to prevail over periods in the future in order to be consistent with the spot-rate structure existing today. For example, suppose the spot rate is 6% for a six-month bond and 7% for a 12-month single-payment bond. An investor who buys the existing six-month bill at 6% and holds it to maturity would have to reinvest the proceeds from that bill in a new six-month bill at a "forward rate" of about 8%, in order to realize the same 7% return that he would have obtained over a one-year horizon by buying the one-year spot instrument in today's market. This forward rate has many interpretations: it could represent a breakeven, an equilibration rate or a market expectation. For shorter maturity bonds, the forward rate structure can also be exploited through arbitrage and through trading in the futures market. Again, what was once a relatively theoretical construct has begun to play a viable and visible role in the market.

Duration

Although yield-to-maturity is a basic measure of long-term value, most market participants must now be concerned with shorter term effects as well. The standard measure of short-term value is simple rate of return: income receipts and accruals are combined with the realized gains (or losses) in principal and expressed as a percentage of the market value of the investment at the beginning of the holding period. For a bond the key uncertainty affecting this rate of return is the potential for a change in price.

Because a price change is equivalent to a change in the yield-to-maturity of the bond, it is useful to plot prices against their associated yields *(see Figure 4, page M18)*. For a bond, the

graph of price against yield is a convex curve with an upward bend. The slope of the line tangent to this curve at the point corresponding to the current interest rate closely approximates the change in price that would be caused by a small change in yield. This slope divided by the current price is the economic elasticity of the bond: the percentage price change per unit instantaneous move in the yield. This elasticity value is called the duration of the bond, and it serves as a valuable gauge of the investment risk associated with the potential volatility of interest rates.

Before the concept of duration was introduced, maturity was the primary benchmark for risk. Longer bonds do tend to have higher durations, but for coupon-bearing bonds, duration saturates with increasing maturity.

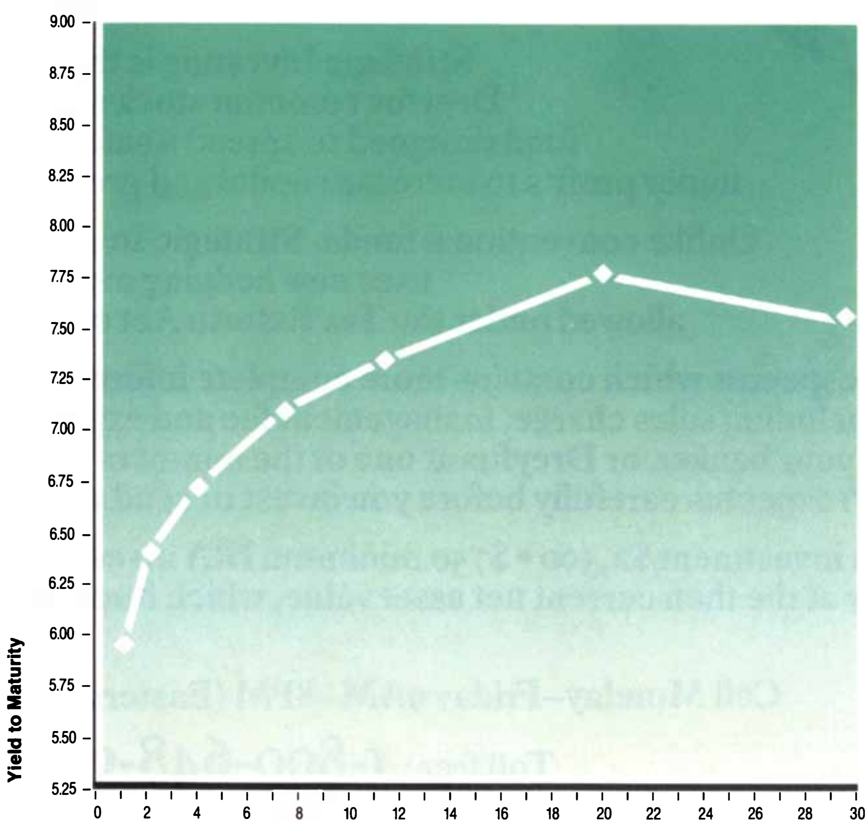
A bond's duration is highly dependent on its coupon rate as well. The duration of a zero-coupon bond whose only cash flow is paid at maturity is essentially equal to the bond's life. Bonds with positive coupon payments have lower durations than a zero-coupon bond of the same maturity. In general, bonds with lower coupons have a higher duration than do higher coupon bonds with smaller maturities.

It is interesting to note that as the maturity of a coupon-bearing bond increases, its duration approaches the duration of a perpetual bond, a bond that has no maturity date. It turns out that the duration of a perpetual depends solely on the interest rate; for example, if the perpetual yields 8%, its duration will be 12.5 years. Since the perpetual has an infinite maturity, the term "duration" has obviously gained a technical connotation far removed from its literal meaning.

The discovery of the concept of duration is an intriguing vignette in the annals of science. It was first proposed by Frederick R. Macaulay in 1938. Macaulay was seeking a more comprehensive and rational discription for the average life of a fixed income security. He concluded that the present-value-weighted average of the time to each of the cash flows would constitute a logical choice for this purpose; the quantity is now known as Macaulay duration.

Macaulay's book was published by the National Bureau of Economic Research and had wide circulation in the U.S. among economists and bond market participants. It was not until 1971, however, that Lawrence Fisher and Roman L. Weil of the University of Chicago observed that the Macaulay duration was closely linked to the price elasticity of a bond. The slope of the price-yield curve multiplied by a simple factor is equal to the Macaulay duration. Thus Macaulay's better definition of average life is directly related to the sensitivity of the price of a bond to interest-rate movements. The intriguing aspect of this observation is that it took 30 years to make. This surely cannot be ranked as a high example of the dissemination of scientific findings throughout the academic world.

Figure 3



U.S. Treasury Yield Curve is a graph of the current yield-to-maturity for Treasury bills, notes and bonds priced at par, plotted against the time until maturity of each issue; the yield curve is shown for February 20, 1987. In most, but by no means all, interest-rate environments yield increases with the time until maturity; one frequent explanation is that investors demand higher yields for tying up funds in issues with a longer time until maturity. The shape of the yield curve is a key factor in determining trading patterns in the securities market.

Portfolio Immunization

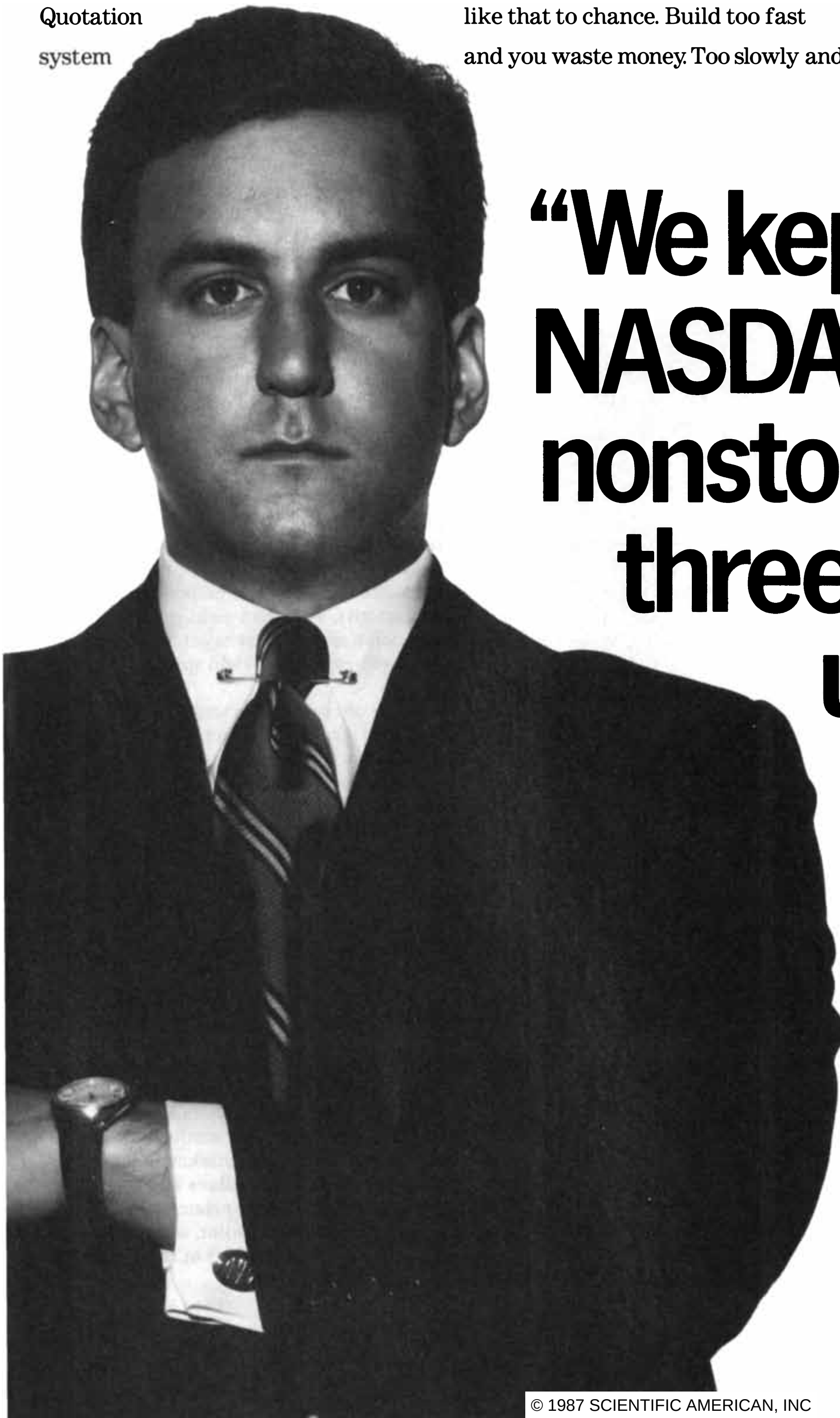
The concept of duration is also central to the important idea of portfolio immunization. The problem is to construct a portfolio of bonds that will be able to satisfy a given schedule of future payouts regardless of future interest-rate movements. It turns out that this general problem reduces to that of finding a bond portfolio whose properties match those of the future liabilities in three respects. First, the present value of the cash flows generated by the portfolio must be equal to the present value of the liabilities at current market interest rates. Second, the duration of the portfolio must match the duration of the liabilities. Third, if one plots the difference between the present value of the portfolio and the present value of the liabilities for a small range in yield both above and below the current interest rate, that difference must attain its minimum value at the current interest rate. If these three conditions are satisfied, the cash flows of the portfolio will meet the future

An electronic stock market. An international network of computers through which shares are electronically traded with incredible speed.

The market is NASDAQ, and the network was built with Unisys equipment. The National Association of Securities Dealers Automatic Quotation system was founded in 1971. By 1978, daily volume was 11 million shares.

Now, in 1987, an average of 155 million shares a day are traded on a network of 3,000 terminals over a 6 million square mile trading floor. Uptime for the central Unisys computer is 99.92%.

"You can't leave network growth like that to chance. Build too fast and you waste money. Too slowly and you compromise service," says Sam Vail. His Unisys team is responsible for helping NASDAQ plan and manage the network growth. "We've been through three generations of equipment without once stopping for software conversions," Sam proudly points out. "Unisys systems grow right alongside the customer. I guess that's what the power of 2 means." Unisys and NASDAQ. The power of ^{2}.



"We kept NASDAQ running nonstop through three major upgrades."

Sam Vail, Account Executive, Unisys.

UNISYS

The power of ^{2}

liabilities at current interest rates and exceed the liabilities if interest rates change by a small amount in either direction. The term "immunization" and the solution described above was first presented by the British actuary F.M. Redington in 1952. He too was quite unaware of Macaulay's definition of duration or with its relevance to the price sensitivity function.

Since the price-yield curve bends upward, there are prices along the curve that lie about the duration tangent line. This curvature above the duration line is somewhat loosely called convexity. There have been several efforts to specify a common benchmark for convexity effects: for example, one proposal is that convexity at any point on the price-yield curve be defined as the average deviation of the curve from the tangent line for yield changes of plus and minus one percent. It turns out, however, that the most severe departures from a linear price-yield relation reflect properties of the bond contract that have nothing to do with the nonlinear terms of the price-yield formula.

Pricing "Off the Curve"

With the increasingly dominant role played by the Treasury yield curve, the yields of virtually all kinds of bonds have come to be stated as yield spreads "off the curve." Such a spread can be understood as the difference between the yield of the security itself and the yield of a Treasury security with a comparable maturity. Thus a 10-year corporate bond would be described as having, say a 75-basis-point (.75%) spread off the curve, that is, a .75% greater yield than a 10-year Treasury bond.

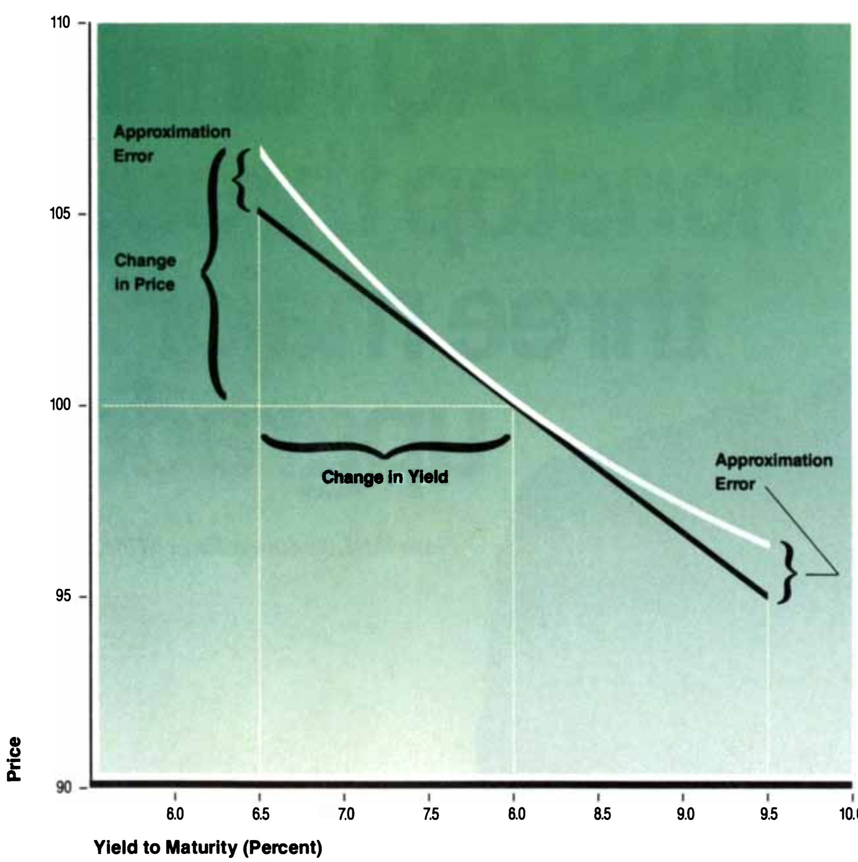
As long as the yield spread remains constant, a given basis-point move in the 10-year Treasury market will generate a corresponding yield move in the corporate security. When the yield spreads themselves change, however, the relation between a bond and its corresponding Treasury can become highly distorted. The price response of the bond can then depart radically from the price response one would associate with a given yield move in the Treasury market.

For example, many kinds of investments, including the typical corporate bond, the mortgage pass-through security and even most long Treasuries, carry some form of "call option." The option allows the borrower, or issuer of the bond, to "call," or repay, the principal of the loan at a specified price before its maturity. The call is a valuable right, and it takes on increasing significance when lower interest rates make it economical for the issuer to exercise the option and refinance the loan. As rates decline, the price of the bond begins to reflect the value of the call option, and the response of price to continued declines in yield becomes increasingly sluggish. In other words, the yield spread relative to the corresponding Treasury security widens to reflect the increasing value of the option embedded in the security. The price-yield relation for such a security traces across bands of increasingly wide yield spread.

The unfavorable price-yield relation that results from such call options departs radically and adversely from that of a straight duration line, and so such behavior is sometimes referred to as negative convexity. With the use of modern option theory it has become possible to characterize many such option effects and thereby develop more refined price functions for various kinds of fixed-income securities.

Mathematical underpinnings are not a new feature of the bond market, but in the past two decades they have become greatly enriched, both in theory and in practice. Nevertheless, one must keep such analyses in perspective. All the current mathematical sophistication enables market analysts to assess only relative values and the correspondingly appropriate yield spreads. Beyond that point, science departs, and the mysteries of the marketplace take over. ◆

Figure 4



Price Yield Curve shows how the price of a bond varies with its yield-to-maturity. The yield-to-maturity is determined by current interest rates; as it increases, price declines, and vice versa. The price volatility of the bond is the change in price for a given change in yield; price elasticity is the percentage price volatility. The illustration shows the price-yield curve for the 8% five-year bond; notice that when the yield-to-maturity is equal to the coupon rate, the price is 100, or par. The tangent line at that point closely approximates the price volatility of the par bond for small changes in yield. For large changes the price-yield curve exhibits convexity, that is, it bends away from the tangent line.


© 1987 Allied-Signal Inc., PO Box 2245R, Morristown, NJ 07960.

THE TERM STRUCTURE OF UNCERTAINTY: ASSESSING INTEREST RATE RISK

◆

by James J. Kennedy and David R. J. Haan
Merrill Lynch Capital Markets

To make the most of their money, Wall Street firms must routinely deal with large potential risks. For example, our firm devotes about $1 billion of its capital to its institutional dealer business. In the course of this business, the firm carries about $30 billion in securities on its books. Were these securities to uniformly fall a mere 3% in value, it could wipe out the firm's capital—a risk that Merrill Lynch cannot afford to take.

Such risks are addressed by careful hedging of the newly issued securities that the firm brings to market and of the inventory necessary to service customer order flow. The hedge position is designed so that a decrease in the market value of these holdings will be offset by a corresponding increase in the market value of the hedge. The hedge securities must be chosen so that they closely compliment the response of the holdings' value to changes in the various factors affecting value. Since the price of any given security may respond differently to changes across these factors, and may in fact have risks uniquely associated with it, the object of hedging is to minimize such risk, not eradicate it. Moreover, the cost of maintaining the hedge must not exceed the profits resulting from the dealer business, otherwise, the potential for loss is merely being replaced by near-certain unprofitability.

In theory, the debt markets should be easier to hedge than the equity markets. Stock prices are more closely linked to the performance and prospects of individual companies than are bond prices, which are more heavily influenced by prevailing interest rates.

At first glance, it might appear that an obligation to buy a particular bond might be adequately hedged by an obligation to sell any other bond. This is not the case, since different bonds react differently to changes in interest rates. Part of our work in the Trading Analysis Group is to determine how individual bonds will respond to changes in interest rates. Our discussion will be confined to Treasury notes and bonds (with maturities of two to thirty years), which best illustrate pure interest rate risk. The starting point for this analysis is the classical equation of bond pricing, stated in either its continuous or discrete form *(see Figure 1, page M22)*.

As these equations show, the price of a bond will depend upon coupon and maturity as well as interest rate levels; each cash flow is discounted by the interest rate applicable to the timing of the cash flow. As interest rates rise, the discounting of these cash flows becomes stronger, so that bond prices fall. But interest rates vary across term; (i.e., today's one-year interest rate may differ from the two-year interest rate). This term structure of interest rates cannot be observed directly, but must be inferred from observed market prices for some number of debt securities.

The lack of direct information about interest rates is traditionally resolved by using yield as a proxy measure for interest rates. Yield is defined to be the internal rate of return for holding a particular bond to maturity, corresponding to a flat term structure for that bond. The response of prices to changes in yield captures much of the interest rate sensitivity; as yields rise, prices fall, and as either coupon or maturity increases, so does the price change for a given yield change. In turn, the most common proxy for term structure is the yield curve, which distributes available yields across maturities.

A number of measures have been created to describe price sensitivity to yield change. One of the most straightforward is dP/dY, which gives the change in price for any small change in yield. This is essentially a linear measure; since the price-yield relationship is in fact nonlinear, another measure, called "convexity", gives the change in dP/dY as yield changes. (In mathematical terms, dP/dY and convexity are the first and second derivatives of the basic bond pricing equation, respectively; they are often used in the Taylor series approximation to the price-yield relationship.)

An important use of dP/dY and convexity is in determining hedging strategies. The simplest of these methods assumes that yield changes are identical for all of the securities under consideration. The ratio between dP/dYs for two securities determines the position in one security which is equivalent to a specified position in the other, (i.e., which will move one-to-one in value for a given yield change). For example, given a long position of $10 million in bond A with dP/dY = -5, an offsetting position in bond B with dP/dY = -4 will require shorting $12.5 million face value of that bond. (Note that dP/dY is less than zero; this arises due to the inverse relationship between price and yield.) Aggregate dP/dY measures may be derived for entire portfolios by weighting individual securities' dP/dY by the corresponding positions.

Since yield changes are not in fact uniform, a better hedging strategy arises from relaxing the assumption that they are. Pure analytical techniques, or regressions against historical yield changes, may suggest a linear relationship between yields for differing securities. The linear coefficient may then be used to fine-tune the

Managing risk is the key to success on Wall Street. Not surprisingly, much of financial theorists' time and effort is taken up with risk: defining it, measuring it, and learning how to avoid it. The key to avoiding risk is learning how the prices of various securities will move relative to each other: what will go up when something else goes down. The task is trickier than first meets the eye.

Master of Possibilities: David L. Wolper.

Producer: Liberty Weekend,
'84 L.A. Olympic Ceremonies,
The Thorn Birds, Roots.

"Substance is the difference between an extravaganza and a historical event."

The '84 Olympics and Liberty Weekend were more than just extravaganzas. They were a celebration of human values, captured in a universal outpouring of pride and spirit. It was the substance of these events that allowed us to not only celebrate history but to become a part of it.

From a more personal perspective, I have found that substance makes the difference in almost everything-even a credit card. That's why I carry a Gold MasterCard® card. Not only can I use it in more places than any other gold card, it provides me with guaranteed hotel reservations, easy cash access and emergency travel services. And with the amount of business travel I do, the substantial credit line has really come in handy more than once.

Whether I'm celebrating my wedding anniversary with my wife, Gloria, or the anniversary of a country with millions, my Gold MasterCard makes anything possible.

Master The Possibilities™

price sensitivities for these securities to arrive at an improved estimate of the hedge ratio. The hedge ratio will thus depend upon the relative movements of yield as well as the response of prices to changes in yield.

This improved hedging technique broadens the selection of eligible hedge instruments. But it carries an important caveat: the more dissimilar the holdings and the hedge, the less correlation exists between yield changes and the more nonlinear are the price changes resulting from given changes in yield.

These issues require another level of sophistication in yield-based hedging techniques. The lion's share of the nonlinearity of the price-yield relationship is captured by the convexity term (d^2P/dY^2). Convexity may be aggregated over a portfolio by the same procedure as for dP/dY, and may be similarly matched against candidate hedge instruments. This method provides protection against a greater range of yield changes, but it relies more heavily upon the assumptions regarding relative yield changes. Furthermore, the difficulties of matching both dP/dY and convexity result in trade-offs between optimizing the hedge with respect to each.

Although useful, there is an inherent difficulty with all of these yield-based hedging techniques. Yield is a summary measure for interest rates; it discards information about the underlying term structure. While historical experience may show good correlation between yield movements, extrapolation to future interest rate environments may not be justified. But while we note the failings of yield-based hedging, we lack sufficient knowledge of the term structure of interest rates to correct it explicitly.

The need for viable hedging instruments which closely track the price performance of the Treasury market is largely responsible for the burgeoning of markets for financial futures and options contracts. Options, futures, and similar instruments, while not strictly debt, derive value from an underlying agreement to buy or sell debt securities at or before expiration of the contract. Since this value depends directly upon the prices of the underlying securities, the use of these contracts to hedge the underlying securities eliminates most elements of interest rate risk. In more general hedging applications, the debt-related contracts may be used as surrogates for the underlying debt securities. The fact that the cost of these contracts is small compared to the underlying securities permits offsetting positions to be easily and cheaply established.

Exchange-traded debt options are particularly important. An option contract confers upon the holder the right, but not the obligation, to buy or sell a specific security at a prespecified price and time; this gives rise to a highly asymmetric price performance for the option relative to price or yield movements in the underlying security. At best, the option price moves one-to-one with (or against) the price of the underlying security; at worst, the option price does not move at all (from near zero value). Thus, interest rate risk can be addressed without sacrificing the possible rewards related to interest rate movement. However, options do incur other types of risks, most notably volatility risk (to which we will return).

Since the price of the underlying security is an important factor affecting option prices, a standard measure of option risk is the amount by which the option price will change for any small change in the price of the security. That measure, called delta, also determines how many options must be held in order to hedge the underlying security. But delta also varies with price; the measure of how delta changes for any small change in price, (i.e., the "price convexity") is called gamma.

Delta and gamma may be combined with yield sensitivity measures for the underlying security in order to reconcile the option's price behavior with yield changes. These measures may then be aggregated across holdings in debt and debt options. But neither delta nor gamma take into account the asymmetry of the option's price performance. Thus, these option-related measures are valid for a much narrower range of price or yield changes than is possible with dP/dY and convexity for debt.

Volatility, measured as the expected standard deviation of prices or yields

Figure 1

Continuous

$$\mathbf{P} = \int_{t_o}^{t_m} C_t \quad D(t)\, dt$$

$$\mathbf{D(t)} = e^{-\int_{t_o}^{t} \hat{r}(t')\, dt'}$$

Discrete

$$\mathbf{P} = \sum_i C_{t_i} \quad D(t_i)$$

$$\mathbf{D(t)} = 1/(1+r_t)^{t-t_0}$$

where:

P = Security price
C_t = Cash flow (coupon or principal) at time t
$D(t)$ = Discounting from time t:

Discrete: r_t = Term interest rate for time t
Continuous: $\hat{r}(t)$ = Forward interest rate(s) through time t

Specification of the term structure of interest rates has centered on the description of the discounting function D(t). The market statement of yield is the value r (=r_t for all t) which solves the discrete form of the equation. Financial theory improves upon this by two distinct approaches: 1) describing the term structure by a single (short- or long-term) rate and a number of shape parameters to capture the functional form, or 2) direct estimation of the term structure by fitting D(t) against observed prices and the associated cash flows for a number of securities. Both methods imply some expected value of D(t) to be imbedded within the integral or sum. One aspect of our current work is to improve upon such methods by determining the proper form for an expectational operator outside the integral or sum.

over time, is also important in determining option prices. If the future price of a security were known, there would be little reason to hedge. Options derive value in part from the magnitude of fluctuations in security prices, rather than any forecast of future price levels. As with interest rates, volatility may not be directly observed; however, volatility may be inferred from option prices through the use of an option model such as the Black-Scholes equation, or approximated by means of historical estimation.

Increases in volatility are reflected by increases in the "time value" imbedded in the price of the option. The Black-Scholes model shows that option price sensitivity to volatility is directly proportional to the option gamma, or price convexity. This has immediate implications for the relative pricing of bonds. Bonds with high convexity tend to have superior price performance for given changes in yield; they appreciate faster in a rising market and fall more slowly in a declining market. They are therefore available at lower yields (higher prices) than low convexity bonds. The differences in these yields are determined by the expected magnitude of yield fluctuations, (i.e., by yield volatility).

Volatility exhibits its own term structure. Price volatility tends to increase with increasing term to maturity, while yield volatility, on the other hand, tends to decrease. In fact, the term structure of volatility is intimately related to the term structure of interest rates, in that observed volatilities provide information about the variance and covariance of interest rates.

In order to use this information, we impose an expectational operator on the right-hand side of the classical equation for bond pricing. This captures the contribution to value due to volatility, and generalizes the formula to bonds which may have uncertain maturities or cash flows. But the theoretical elegance of the revised bond pricing formula is offset by the practical difficulties of accurately determining its specific form.

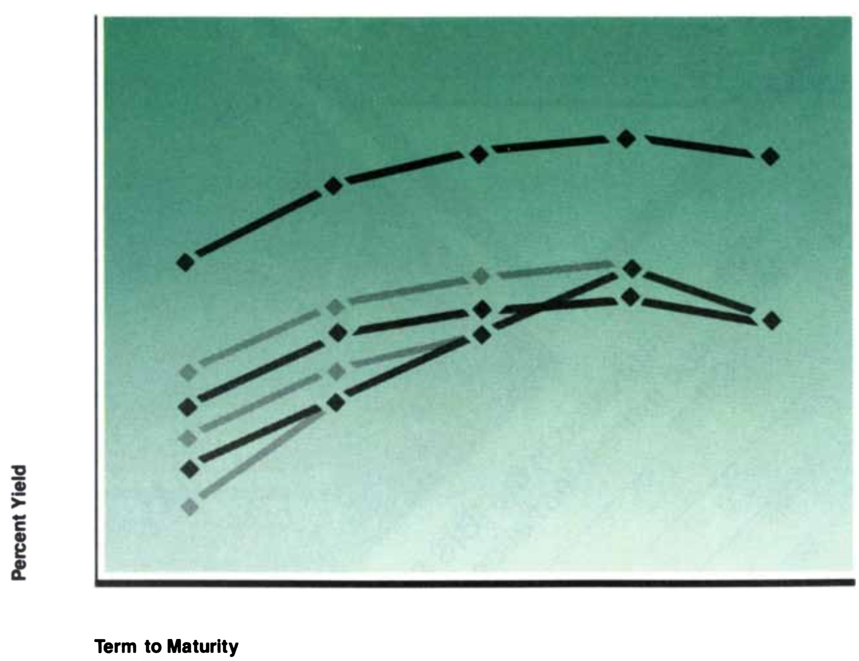
Here at Merrill Lynch, we are currently working to refine a model which simplifies many of the practical considerations of analyzing the revised bond pricing formula without sacrificing its generality. Specifically, we constrain the expectational operator to take a tractable form. We reduce the covariance to two variance measures: the first, "parallel" volatility, measures yield shifts across the entire yield curve, (i.e., across bonds of all maturities), while the second, "adjacent" volatility, measures the smaller magnitude yield shifts between adjacent maturities.

The mechanics of this model may be visualized as a multiply segmented pipe representing the yield curve *(see Figure 2, page M23)*. Each maturity is represented by a joint in the pipe, which may be either rigid or loose. Starting with all joints rigid, the entire structure is moved up or down some amount (parallel shift). One end of the structure is then fixed. The joints for the segment at that end are loosened, and the free end of the structure is moved slightly (adjacent shift). The segment at the end is then fixed in place. This operation is repeated sequentially for each segment of the structure until every segment is fixed. The result is the full perturbation of the yield curve.

This simple model gives rise to a covariance structure in close agreement with historical experience of changes in the yield curve for current Treasury issues. A smoothing function of small weight damps out anomolous yield structures, such as zig-zags, which do not normally arise in the markets. With this additional feature, the model captures a great deal of the behavior exhibited by the diffusing yield curve. This model appears to apply equally well to the term structure of interest rates.

This perspective of covariance of the term structure of interest rates increases our ability to quantify changes in interest rates, and to relate these to risk for debt securities. Though incomplete without a full specification of the term structure, this perspective permits us to make assertions about the way that changes in composite measures take form, and refine our analytical methods for assessing and minimizing interest rate risks accordingly. This in turn permits Merrill Lynch to better control not only its own risk, but also the risk of portfolios that the firm manages for its clients, and to truly make the most of its own, and its customers', money. ◆

Figure 2



A simple model accounts for the covariance structure of the yield curve. Starting with an initial shape (heavy black) a parallel shift is applied to the entire curve (light black). Then one end of the curve is fixed, and the remainder of the curve is shifted in relation to this end (light gray), fixing the value for the maturity adjacent to the one previously fixed. This operation is repeated for each segment of the yield curve, until values for every maturity have been fixed; this gives rise to a new instance of the yield curve (heavy gray). This gives rise to the same covariance structure for yields as has been experienced historically. With minimal smoothing, this model realistically captures the richness of yield curve diffusion.

WE BELIEVE THAT THE GENIUS OF THE FUTURE LIES NOT IN TECHNOLOGY

The Anatomy of Memory

An inquiry into the roots of human amnesia has shown how deep structures in the brain may interact with perceptual pathways in outer brain layers to transform sensory stimuli into memories

by Mortimer Mishkin and Tim Appenzeller

Within the small volume of the human brain there is a system of memory powerful enough to capture the image of a face in a single encounter, ample enough to store the experiences of a lifetime and so versatile that the memory of a scene can summon associated recollections of sights, sounds, smells, tastes, tactile sensations and emotions. How does this memory system work? Even defining memory is a struggle; introspection suggests a difference between knowing a face or a poem and knowing a skill such as typing. Moreover, the physical substrate of memory, the 100 billion or so nerve cells in the brain and their matted interconnections, is fantastically intricate. But in a tentative and schematic way, my colleagues and I (Mishkin) can begin to describe how the brain remembers.

The picture we have arrived at is largely anatomical. Over the past 20 years we have identified neural structures and stations (large arrays of cells) that contribute to memory, traced their connections and tried to determine how they interact as a memory is stored, retrieved or linked with other experience. Other investigators analyze memory on a finer scale: in some of the simplest animals and in neural tissue isolated from higher animals they have detected changes in the electrical and chemical properties of single neurons as a result of simple kinds of learning. The complexity of our subject, memory in the human brain or—as a best approximation—the brain of Old World monkeys, calls for a different initial approach, one emphasizing broad-scale architecture. Ultimately, to be sure, memory is a series of molecular events. What we chart is the territory within which those events take place.

Many of the studies leading to the current picture were suggested by case histories of patients who, because of disease, injury or surgery affecting specific areas of the brain, lost some of their ability to learn or remember. Probably the most famous of these cases is a profoundly amnesic patient known as H.M. As the subject of studies by Brenda Milner of the Montreal Neurological Institute and her colleagues from other institutions, H.M. has provided an abundance of information about the pattern of impairment associated with a specific kind of brain damage.

The experimental work, done mostly in macaque monkeys, has combined anatomical, physiological and behavioral investigations. Tracers that are carried along axons, the slender processes through which neurons send signals, have revealed the neural circuitry that might enable specific structures to play a role in memory. Measurements of the electrical activity of neurons or their uptake of radioactive glucose have distinguished parts of the brain that are active during tasks related to learning. The final category of experiment, meant to assess the functional importance of structures identified by other means, has combined surgery or drug administration with psychological testing. In the brain of an experimental animal, stations are destroyed or blocked by drugs, or the pathways linking them are severed. The animal is then examined on behavioral tests meant to tease apart the various components of memory and determine which of them is impaired.

Our route to understanding human memory is an indirect one, with unavoidable drawbacks. The macaque brain is about one-fourth the size of the brain of the chimpanzee, the nearest relative of human beings, and the chimpanzee brain in turn is only about one-fourth the size of the human brain. With the increase in size has come greater complexity. The structures we study in the macaque all have counterparts in the human brain, but their functions may well have diverged in the course of evolution. The unique human capacity for language, in particular, and the cerebral specializations it has brought set limits to the comparative approach. Yet basic neural systems are likely to be common to monkeys and human beings, and our findings have been consistent with what is known directly about human memory loss.

The Visual System

Most often memories originate as sensory impressions. Before one asks how the brain stores a sensory experience as a memory, one would like to know how the brain processes sensory information to begin with. A study of the neural pathway responsible for visual perception was in fact the starting point for our inquiry into memory.

The central visual system begins at the striate cortex, or primary visual cortex, an area on the back surface of the brain that receives information about the visual world from the retina, by way of the optic nerve and an intermediate station (the lateral geniculate body) deep in the brain. The striate cortex registers a systematic map of the visual field: each small region of the field activates a distinct cluster of neurons. The visual system does not end in the striate cortex, however. By the 1950's it was clear that the temporal lobe, the division of each brain hemisphere lying behind the ear and temple, also has a role in vision.

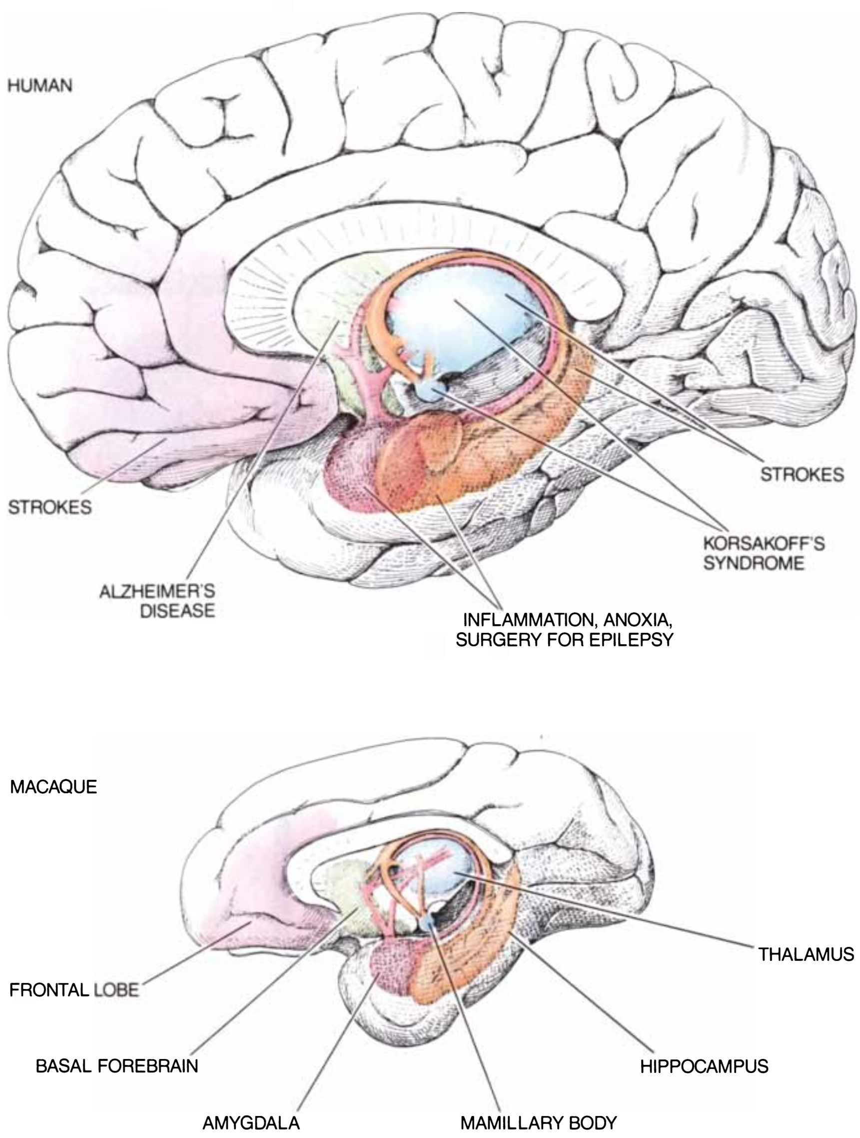
The visual area in the temporal lobe, other workers and I learned in the 1960's, is in fact the continuation of a pathway that begins in the striate cortex. The pathway extends forward through cortical tissue (the outer layers of the brain) into the inferior temporal cortex, on the lower surface of the temporal lobe. Neuroanatomical studies showed that a number of distinct cortical stations are connected in various sequences along the pathway.

VISUALLY DISTINCTIVE OBJECTS serve for testing memory in monkeys. The author (Mishkin) and his colleagues have sought to identify the structures and pathways in the brain that enable a monkey to recognize the familiar object in a pair, having seen the object just once before. Using a plethora of objects ensures that the monkey must learn a totally new sample in each trial.

Investigators in several laboratories explored the contribution of particular stations to visual perception by surgically damaging the pathway in monkeys and then testing the animals on visual tasks, and also by recording electrical activity from each station in animals exposed to various visual stimuli. In a crucial experiment, Charles G. Gross of Princeton University and his colleagues recorded the responses of neurons in the inferior temporal cortex to stimuli—small shapes—displayed to the monkeys.

It was already known that individual neurons in the striate cortex respond most strongly to a simple stimulus, such as a short line with a specific orientation, presented at a specific location in the visual field (a phenomenon discovered by David H. Hubel and Torsten N. Wiesel). Inferior-temporal neurons recorded by Gross and his colleagues, however, responded to more complex shapes within an area averaging 20 to 30 degrees on a side. Certain neurons even responded to a complex shape wherever it was placed in the visual field. The findings suggested each inferior-temporal neuron receives data from large segments of the visual world, and often about the entire constellation of properties making up a visual stimulus.

These results and others led us to postulate that visual information is processed sequentially along the path. Neurons in the pathway have "windows" on the visual world that become progressively broader, in both their spatial extent and the complexity of the information they admit, at successive stations. The cells respond to progressively more of an object's physical properties—including its size, shape, color and texture—until, in the final stations of the inferior temporal cortex, they synthesize a complete representation of the object.



SITES OF BRAIN DAMAGE by diseases and other events that can cause memory loss in human beings are revealed in a cutaway of the human brain (*top*). Color coding indicates the corresponding structures in a macaque brain (*bottom*), where they are labeled. The human brain is shown about two-thirds its normal size; the monkey brain is life-size.

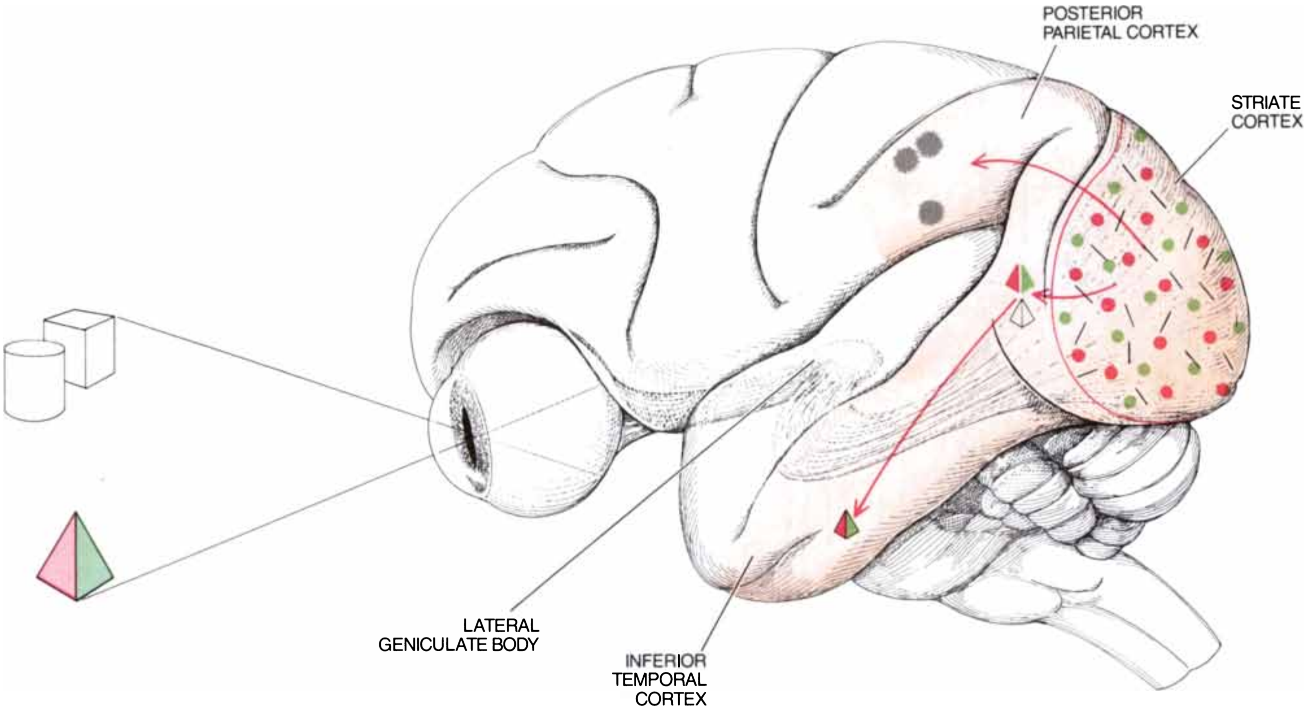
From Experience to Memory

Along the visual pathway, then, the brain integrates sensory data into a perceptual experience. Data from other senses seem to be processed in much the same way; in our laboratory at the National Institute of Mental Health, David P. Friedman, Elisabeth A. Murray, Timothy P. Pons and Richard J. Schneider recently traced an extended processing pathway for tactile sensations. At the first station individual neurons respond to single points on the surface of the body, whereas neurons in the final station respond to stimuli over broad areas and perhaps to the complete tactile experience.

What more must happen for such integrated perceptions to be stored as memory? The question is still very much under investigation, but the crucial pathways can now be discerned. They are anchored in two structures on the inner surface of the temporal lobe in both hemispheres of the brain: the hippocampus, after the Greek word for sea horse, which the structure resembles, and the amygdala, after the Greek for almond.

Studies of human patients had already strongly suggested that the hippocampus plays a crucial role in memory. Since the 1950's the surgical removal of part of the temporal lobe has been the treatment of last resort for patients with severe epileptic seizures focused in that part of the brain. In the early days of the treatment a handful of patients developed severe amnesia after the surgery; H.M.'s handicap originated in that way. Two features of the memory loss were notable. First, the amnesia was global, extending to memory for experience in all the senses, and second, it was anterograde in nature, in that even though patients retained memories laid down some time before the surgery, they could form no new memories. In every case of such memory loss the surgery had damaged the hippocampus.

Yet it proved impossible to reproduce a comparable, global loss of memory in animals by removing the hippocampus alone. Through a series of experiments focusing instead on the amygdala, we found that it plays as large a part in memory as the hippocampus. By damaging both structures at once, in both brain hemispheres of test monkeys, we arrived at an animal



VISUAL SYSTEM processes information along two pathways in the cortex, the outer layer of the brain. Initial processing of the information (which arrives from the retina by way of the lateral geniculate body) takes place at the beginning of the pathways, in the striate cortex. Individual neurons there respond to simple, spatially limited elements in the visual field, such as edges and spots of color. Along the lower pathway (which in fact consists of a number of diverging and reconverging channels) neurons analyze broader properties of an object, such as overall shape or color. At the far end of this "object" pathway, in the inferior temporal cortex, individual neurons are sensitive to a variety of properties and a broad expanse of the visual world, which suggests that fully processed information about an object converges there. Along the upper cortical pathway, which has not been studied in the same detail, the spatial relations of a scene are analyzed. A perception of an object's position with respect to other landmarks in the visual field, for example, would take shape in the "spatial" pathway's final station, which is situated in the posterior parietal cortex.

model of global anterograde amnesia.

We were motivated at first by a wish to understand just why surgical damage to the inferior temporal cortex in monkeys leaves the animals incapable of a particular kind of visual learning: choosing an object or a pattern that has consistently been associated with a reward of food over one that carries no reward. We believed the difficulty was the result of impaired visual perception, reflecting damage to the highest processing station in the visual system. Conceivably, however, the impairment instead reflected an inability to associate a stimulus with a reward.

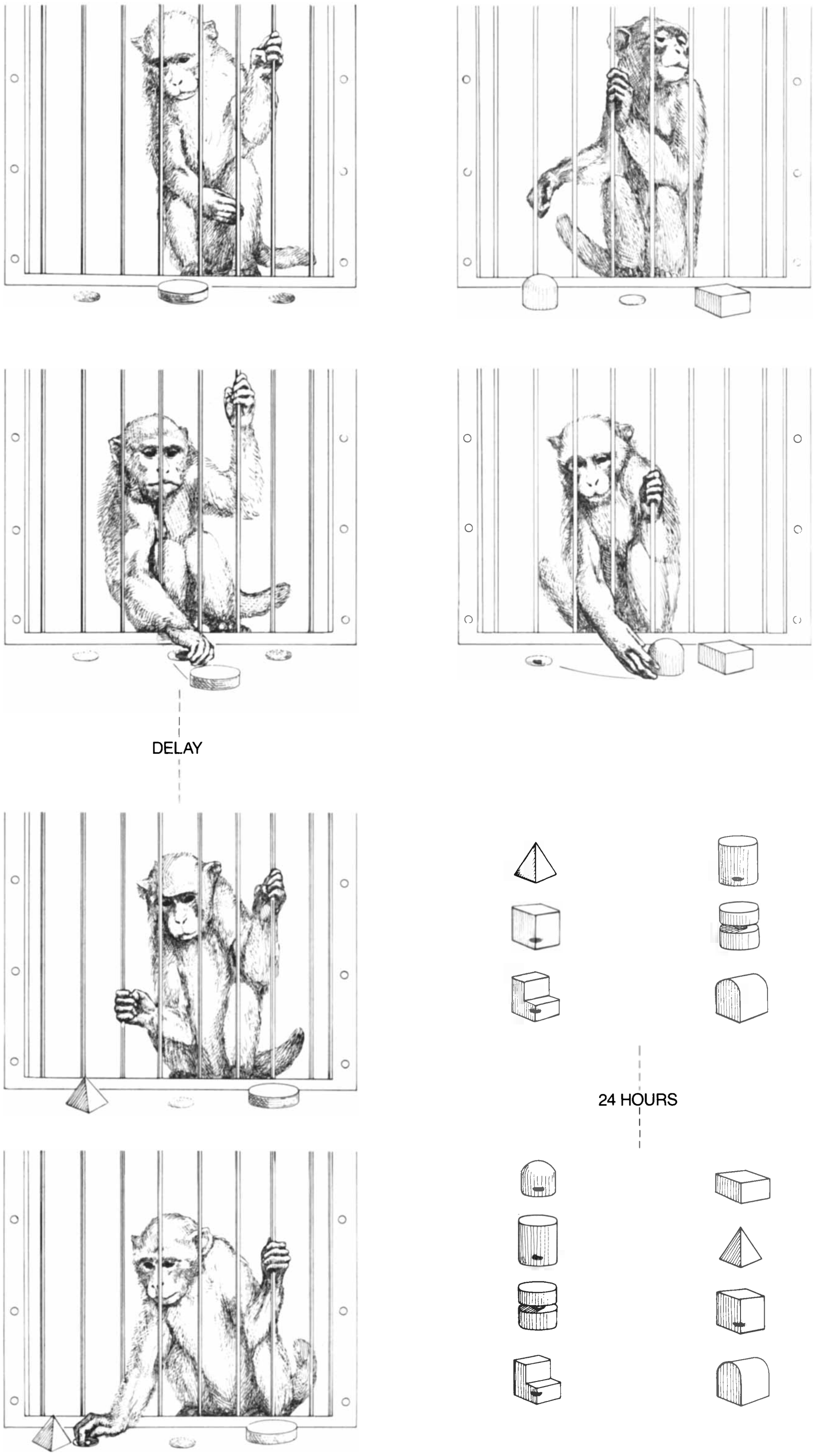
One way to rule out the latter possibility, and thereby confirm that the difficulty of monkeys with inferior-temporal lesions was a visual one, would be to show that a different structure, undamaged in the test monkeys, is responsible for attaching reward value to a visual stimulus. With Barry Jones, then at McMaster University in Ontario, I looked for structures with which the visual system has anatomical links and tested whether surgically destroying those structures would affect the ability of monkeys to choose a baited object. Two structures we tested were the amygdala and the hippocampus, which have extensive connections (indirect ones in the case of the hippocampus) with the inferior temporal cortex.

The poor performance that resulted from bilateral removal of the amygdala suggested that it is the structure largely responsible for adding a positive association—the expectation of a food tidbit—to a stimulus processed by the visual system. Before studying the interaction of the visual system and the amygdala in greater depth, Brenda J. Spiegler, then at the University of Maryland at College Park, and I sought a way to increase the degree of impairment. We expanded the surgery to include both the amygdala and the hippocampus, whose removal by itself had had no effect.

Animals whose amygdala alone had been removed were slow to learn the association of stimulus and reward, but they were still able to do so. We were therefore startled to find that removing the amygdala and the hippocampus in combination did away altogether with the monkeys' ability to perform the task. The dramatic increase in impairment led us to wonder whether we might now be seeing a deficit that went beyond an inability to associate a familiar object with a reward. Could these monkeys have failed to choose the baited object because they could not remember the object itself from one trial to the next? That is to say, in destroying the hippocampus together with the amygdala, had we created a visual amnesia?

As it happened, Jean Delacour of the University of Paris and I had recently developed a test that was specifically sensitive to visual memory, as distinct from the ability to link an object with a reward. In that memory test, called delayed nonmatching-to-sample, the animal is presented with a distinctive object, under which it finds a reward of a peanut or a banana pellet. Next the animal is confronted with two objects, one of them the object seen earlier and the other an unfamiliar object. The food is now concealed under the new object; the monkey is therefore rewarded for recognizing and avoiding the familiar object in favor of the novel one. Each trial makes use of a totally new pair of objects.

It is a routine that monkeys learn readily. In an adaptation of an approach developed by David Gaffan of the University of Oxford, an animal's visual memory can then be taxed by increasing the delay between the initial display and the choice or by giving the animal not one object but a series of



TWO TESTS OF LEARNING may gauge fundamentally different abilities. In the test known as delayed nonmatching-to-sample (*left*) a monkey is confronted with an unfamiliar object, which it displaces to find a reward. After a delay the animal sees the same object paired with a new one. The monkey must recognize the object it saw earlier and move the new one instead in order to get a reward. When a monkey has mastered the routine, the task can be made more difficult (by increasing the delay, for example) to tax the animal's ability to remember an object seen before just once. A second test, known as object discrimination (*right*), employs a sequence of 20 pairs of objects. One object in each pair conceals a reward. A monkey is shown the same sequence repeatedly, at 24-hour intervals, until it learns to choose the baited object in each pair consistently. The author and his colleagues have shown that brain structures crucial to a monkey's success on the first task can be damaged without impairing its ability to learn the second task.

objects to remember, followed by a series of choices. Because the food is always associated with a novel object, the ability to link a specific object to a reward has no bearing on performance. The reward serves merely as an incentive; the test measures recognition memory specifically. Normal monkeys attain almost perfect scores.

We had already applied this powerful test of visual recognition to monkeys whose inferior temporal cortex had been removed and found that the test registered the same failure to perceive or identify familiar objects that we had observed in earlier tests. We now tested animals whose visual system was intact but whose hippocampus and amygdala had been removed bilaterally. When the delay between the first object and the choice was very short, the animals could perform the task, which suggested that they suffered no defect in visual perception. When we increased the delay to a minute or two, on the other hand, their scores fell nearly to the level of chance. It seemed, then, that we had created a true memory loss.

What is more, the amnesia resulting from combined hippocampal and amygdalar lesions was not confined to visual stimuli; it seemed to be global. In our laboratory Murray found a similar impairment in the test monkeys on a nonmatching-to-sample test of tactile recognition, that is, the ability to recognize objects by touch. In human patients such global amnesia had been thought to result from damage to the hippocampus alone, and in fact a recent postmortem study of one amnesic patient by Larry R. Squire of the University of California at San Diego and his colleagues found lesions in the hippocampus only. Many other amnesic patients have more extensive lesions, however, and damage to the amygdala may have contributed to their impairment. Indeed, an early study of amnesics suggested that the severity of their memory loss might vary in proportion to the amount of damage sustained jointly by the amygdala and hippocampus.

When we went on, with Richard C. Saunders, to examine each structure's contribution to memory in the monkey, we found just such a graded effect. In mediating visual recognition memory the amygdala and the hippocampus seem to be coequal; removing either structure by itself has only a small effect on an animal's recognition ability, presumably because each one can substitute for the other. Removing one of the structures from both hemispheres and the other from only one hemisphere yields much greater impairment. Bilateral removal of both the amygdala and the hippocampus, finally, results in an animal whose score on delayed nonmatching-to-sample is little better than chance.

Further Stations

Damage to the amygdala and the hippocampus, two major components of what is known as the limbic system, is not the only kind of neuropathology that can result in global amnesia. In other amnesic patients the site of the damage is the diencephalon, a cluster of nuclei at the center of the brain that is organized into two structures known as the thalamus and the hypothalamus. Parts of the diencephalon situated medially (near the midline of the brain) degenerate in Korsakoff's syndrome, a global amnesia seen in some chronic alcoholics; diencephalic damage from strokes, injuries, infections and tumors can cause the same amnesic syndrome. The clinical evidence that implicates diencephalic nuclei in memory is reinforced by the anatomical finding that the diencephalon receives fibers running from the hippocampus and the amygdala.

To test the possibility that the diencephalon interacts with the limbic structures in a kind of memory circuit, we again applied our experimental strategy of surgery followed by behavioral testing. John P. Aggleton of the University of Durham and I destroyed, at first in combination and then separately, the regions of the diencephalon to which the hippocampus and amygdala send fibers. Testing for visual recognition ability showed the same pattern of memory failure that had resulted when the hippocampus and amygdala were themselves removed. Combined damage to the diencephalic targets of both structures severely impaired the monkeys' recognition memory, whereas damage to either target alone had only a slight effect. It appeared we had identified two distinct memory circuits, either of which suffices for visual recognition.

That the diencephalon and the limbic structures participate in a circuit rather than making totally independent contributions to memory was confirmed by further studies. In our laboratory Jocelyne H. Bachevalier and John K. Parkinson found that cutting the connections between the structures produced the same memory impairments seen when the structures themselves were damaged.

Earlier neuroanatomical findings suggested that we had not yet followed the circuits to their end. Nuclei in the thalamus that communicate with the limbic structures send fibers in turn to the ventromedial prefrontal cortex: an area of cortex tucked under the front of the brain. There too Bachevalier found that surgical lesions led to a profound loss of recognition memory.

Thus the final station in the visual system, and in the other sensory systems as well, is linked with two parallel memory circuits including at a minimum the limbic structures of the temporal lobe, medial parts of the diencephalon and the ventromedial prefrontal cortex. How do these structures work together in the process of memory? The question is complicated by the fact that memories are probably not stored exclusively or even mainly in the circuits themselves. The clinical observation that damage to the limbic system in humans leaves old memories intact and accessible means they must also be stored at an earlier station in the neural paths we mapped. The likeliest repositories of memory, in fact, are the same areas of cortex where sensory impressions take shape.

The subcortical memory circuits must therefore engage in a kind of feedback with the cortex. After a processed sensory stimulus activates the amygdala and hippocampus, the memory circuits must play back on the sensory area. That feedback presumably strengthens and so perhaps stores the neural representation of the sensory event that has just taken place. The neural representation itself probably takes the form of an assembly of many neurons, interconnected in a particular way. As a result of feedback from the memory circuits, synapses (junctions between nerve cells) in the neural assembly might undergo changes that would preserve the connectional pattern and transform the perception into a durable memory. Recognition would take place later, when the neural assembly is reactivated by the same sensory event that formed it.

Just how each structure in the memory circuits might contribute to the feedback is not known. There are already clues to the nature of the feedback as a whole, however. One clue lies in yet another structure that our work has implicated in recognition memory. It is the basal forebrain cholinergic system, a cluster of neurons that provides the cortex and the limbic system with their major input of a neurotransmitter (a chemical messenger that carries signals across synapses) called acetylcholine.

Acetylcholine seems to play a vital role in memory. For one thing, it is depleted in Alzheimer's disease, one hallmark of which is memory loss. In our laboratory Thomas G. Aigner has found, moreover, that monkeys perform better than they normally do on the test of visual recognition memory when they are given physostigmine, a drug that enhances the action of acetylcholine. When they are given scopolamine, on the other hand, a substance that blocks the action of the neurotransmitter, their performance is impaired. Recently, in collaboration with a team of investigators headed by Donald L. Price and Mahlon R. DeLong at the Johns Hopkins University School of Medicine, Aigner and I also established that damaging the basal forebrain impairs recognition memory in monkeys, although the effect seen so far is not as severe or long-lasting as the effects of damaging the other structures we have studied.

The circuitry that would enable the other structures to enlist the basal forebrain in memory formation is certainly present. For example, the hippocampus and amygdala have extensive projections to the basal forebrain, which in turn sends acetylcholine-containing fibers back not only to the limbic structures but also to the cortex. In a plausible scenario for memory formation, the activation of the subcortical memory circuits by a sensory stimulus would trigger the release of acetylcholine from the basal forebrain into the sensory area. The acetylcholine (and probably other neurotransmitters whose release is triggered in the same way) would initiate a series of cellular steps that would modify synapses in sensory tissue, strengthening neural connections and transforming the sensory perception into a physical memory trace.

Results of a recent biochemical study suggest that a possible mechanism of synaptic modification is active in the area we believe is a likely site of memory storage: the final stations of the visual system. Aryeh Routtenberg of Northwestern University has proposed that the addition of a phosphate group to a brain protein known as *F*1 by an enzyme, protein kinase *C*, underlies the synaptic changes seen after repeated electrical stimulation of certain neurons. To gauge the activity of the phosphorylation mechanism in the monkey visual system, Routtenberg and his student Robert B. Nelson added radioactive phosphorus to tissue that my group had dissected from visual areas. The largest amounts of the tracer were incorporated into *F*1 in tissue from the final stations. The finding may indicate that tissue there has a special capacity to undergo synaptic changes that could store memories.

Types of Memory

The brain architecture described so far was revealed by its contribution to one kind of memory: recognition memory, the remarkable faculty by which a monkey, having seen or felt a distinctive object just once, can recognize the object many minutes later and avoid it in favor of a novel one. There are more complex kinds of memory, of course, some of which can also be tested in monkeys. Their exploration has revealed interesting specializations in the neural paths we traced.

In perceiving an object, for example, one learns not only its distinguishing features but also its location with respect to other objects or landmarks. Remembering a sculpture seems intuitively to be a different task from remembering its position on a gallery floor. Neuroanatomically the task is different as well. To begin with, spatial vision—the ability to see spatial relations—depends on a branch of the visual system different from the one responsible for perception of an object's distinctive qualities.

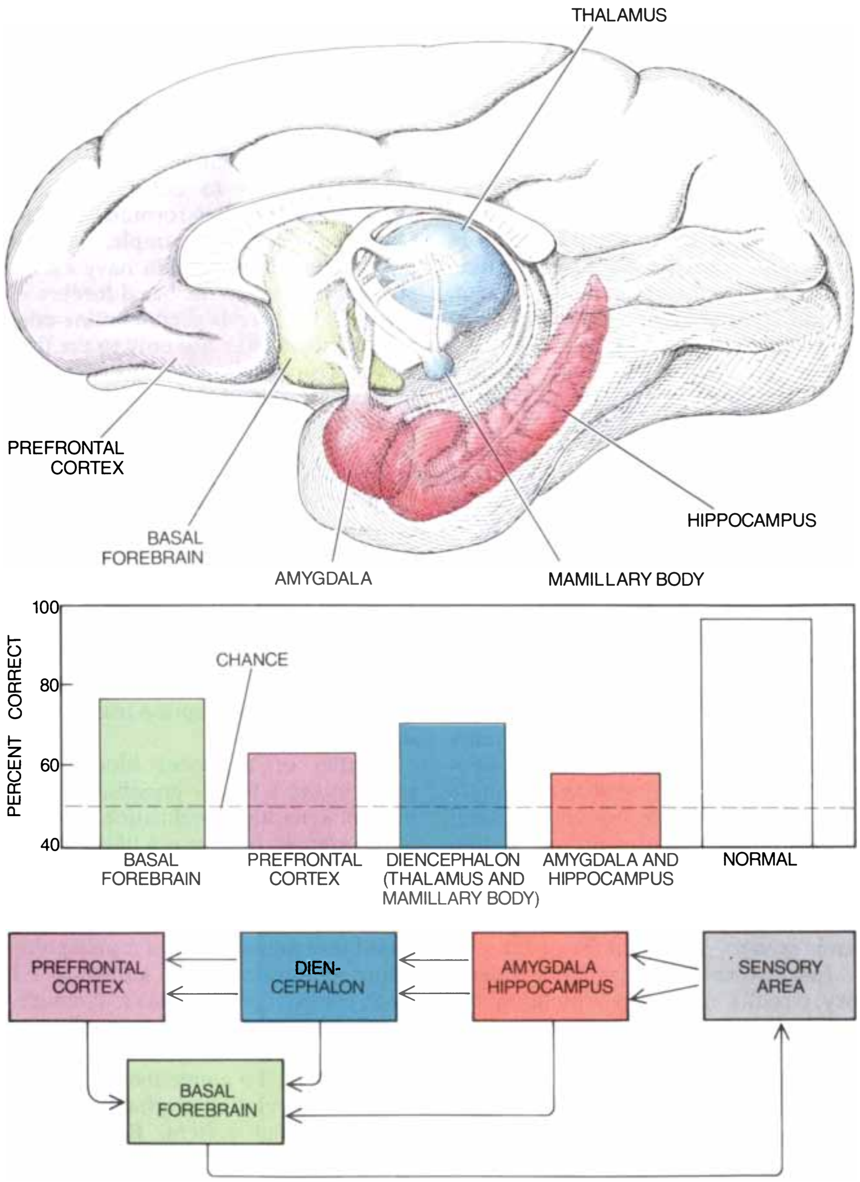
In 1973, working in our laboratory, Walter Pohl confirmed earlier suggestions that tissue in the parietal cortex, near the top of the brain, has a visual role. He showed that its removal leads to a visual impairment—but a very different impairment from the one that follows damage to the inferior temporal cortex. Unlike monkeys with inferior-temporal lesions, animals with parietal damage could still tell distinctive objects apart, but they could not perceive spatial relations.

Pohl's evidence came from a test in which monkeys were presented with two covered food wells. A cylindrical object stood between the wells, closer to one well than to the other in a position that varied from trial to trial. The monkeys were rewarded for uncovering the well nearer the cylinder; it contained a peanut, whereas the other well was empty. The task is relatively easy for animals with inferior-temporal lesions. Animals with damage to the posterior parietal cortex, however, had great difficulty learning to pick the baited well.

Other work, including a metabolic study done by Kathleen A. Macko, Charlene D. Jarvis and me in collaboration with a team headed by Charles Kennedy and Louis Sokoloff, also of the National Institute of Mental Health, confirmed that the posterior parietal cortex belongs to the visual system. We injected a radioactive analogue of glucose and studied its uptake in tasks related to vision. The results pointed to the involvement not only of the inferior temporal cortex but also of posterior parietal tissue.

Furthermore, in our laboratory Leslie G. Ungerleider recently found that another anatomical pathway, in addition to the one we now know is devoted to processing the visual characteristics of an object, emerges from the striate cortex (the primary visual station at the back of the brain). Instead of coursing forward into the inferior temporal cortex, this second pathway runs upward through a series of stations to a final station in the posterior parietal cortex. Spatial relations are probably analyzed along this pathway; from its final station processed spatial perceptions presumably activate the subcortical memory system.

The roles of the two memory circuits outlined earlier may differ in spatial learning. Although the hippocampus and the amygdala can substitute for each other in learning to recognize an object, the former structure seems to be particularly important for learn-

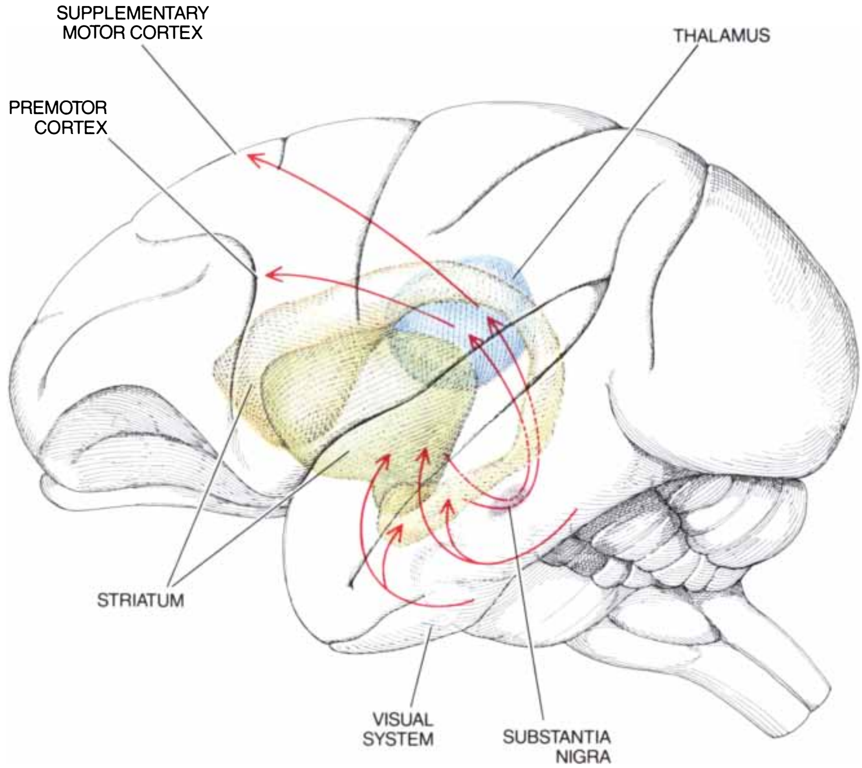


MEMORY SYSTEM was mapped largely by examining monkeys on a test measuring visual recognition memory (*see illustration at left on page 84*) after surgical damage to specific structures or pathways in the brain. The diagram at the top indicates structures that were found to be crucial. The graph at the middle shows the average scores monkeys achieved on the test following surgical damage to specific elements of the system; for comparison, it also shows the nearly perfect performance of normal monkeys. The chart at the bottom (based on a variety of evidence, including neuroanatomical studies of the circuitry linking the structures) shows how the structures might interact in the formation of a memory. A perception formed in the final station of a cortical sensory system activates two parallel circuits. One circuit is rooted in the amygdala and the other in the hippocampus; both encompass parts of the diencephalon and the prefrontal cortex. Each structure in turn sends signals to the basal forebrain. Through its many connections to the cortex, the basal forebrain could close the loop. It might precipitate changes in neurons of the sensory area, which could cause the perception to be stored there as memory.

ing spatial relations. Stimulated by the work of other investigators, whose results had suggested the importance of the hippocampus for spatial learning in rodents, Parkinson explored its role in monkeys.

He trained normal monkeys to perform a test that taxed their memory for the location of objects. In each trial the animals were shown two completely new objects in specific locations on the test tray; next they saw one of the original objects in its original position and an exact duplicate, either where the second object had been or at yet a third position on the tray. The animals were rewarded for choosing the original object in its original position.

After surgery in which their amygdala was removed bilaterally, animals quickly relearned the task and did it accurately. Bilateral removal of the hippocampus, however, left monkeys unable to perform the feat at all. Mary Lou Smith of the Montreal Neurological Institute, working with Milner, recently reported a parallel finding in human amnesics: a correlation between the extent of hippocampal damage and the degree of impairment in remembering the locations of objects.



HABIT FORMATION, or the development of automatic connections between a stimulus and a response, may depend on the striatum. It receives extensive connections from sensory systems in the cortex (here typified by one branch of the visual system) and in turn sends fibers to cerebral structures that communicate with the premotor cortex and the supplementary motor cortex. The neuroanatomy thus provides a relatively direct path through which a stimulus registered in a sensory area could lead to a motor response. Indeed, damage to the striatum or to connections between the cortex and the striatum has been found to hamper monkeys' performance on a visual test measuring habit formation.

Where Memory Meets Memory

The amygdala also has its specializations. Many of the amygdala's distinctive contributions to memory were suggested by its remarkable neuroanatomy, which was known long before its role in memory had been established. The amygdala—or, more precisely, the amygdaloid complex, which consists of several nuclei—is a kind of crossroads in the brain. Many investigators, including Blair H. Turner of the Howard University College of Medicine and me, had shown the amygdala has direct and extensive connections with all the sensory systems in the cortex. It also communicates with the thalamus, along a path that is a segment of the memory system. Finally, the same parts of the amygdala on which sensory inputs converge send fibers deeper into the brain to the hypothalamus, which is thought to be the source of emotional responses.

The number and variety of connections between sensory areas and the amygdala led Murray and me to wonder whether it might be responsible for associating memories formed in different senses. Until we turned to that question we had always investigated learning as it is manifested in recognition, that is, a response to the original visual, tactile or spatial experience. Quite often, however, what awakens a memory is a sensory experience of a different kind. The sound of a familiar voice on the telephone summons a visual memory of the caller's face; the sight of a purple plum brings to mind its taste. Some kind of interchange between the cortical areas where memories in each sense are stored would seem to be necessary for such cross-modal recall. Might it be mediated by the amygdala?

To test the possibility, Murray and I combined visual and tactile versions of the recognition-memory test. We taught monkeys to perform delayed nonmatching-to-sample with objects chosen from a pool of 40, each of them distinctive both visually and tactilely. The animals did the tasks both by sight and in the dark, where they had to rely on their sense of touch to distinguish the sample object from the new one. Monkeys performed both tasks nearly as well after bilateral removal of the amygdala as they had before the surgery. Their visual and tactile recognition memory had remained largely intact, in keeping with our earlier finding that the hippocampus and the amygdala can substitute for each other in mediating recognition memory.

Now that the monkeys were thoroughly familiar with the visual and tactile qualities of the 40 objects, we changed the nature of the test. In each trial the monkey first examined an object in the dark, by touch, but then confronted the same object and an alternative in the light and had to choose between them by sight alone. To recognize the object it had felt a few seconds earlier, the animal had to associate visual and tactile memories. A control group of animals whose hippocampus had been removed performed the task well, choosing the correct object about 90 percent of the time. Animals lacking their amygdala, on the other hand, did little better than they would have done by chance.

The evidence that the amygdala mediates the association of memories formed through different senses may shed light on an old mystery of neuropsychology. Nearly 50 years ago Heinrich Klüver and P. C. Bucy were struck by the strange behavior of monkeys whose temporal lobes had been removed. The animals often examined an inedible object repeatedly and indiscriminately, by touch, taste and smell, as if they found it perennially unfamiliar. Later the same bizarre effect was shown to follow removal of the amygdala alone. Our result led us to propose that one root of the behavior is the monkeys' inability to link different kinds of memory. Seeing a familiar object, they cannot recall how it smells; after smelling it they still cannot recall its taste.

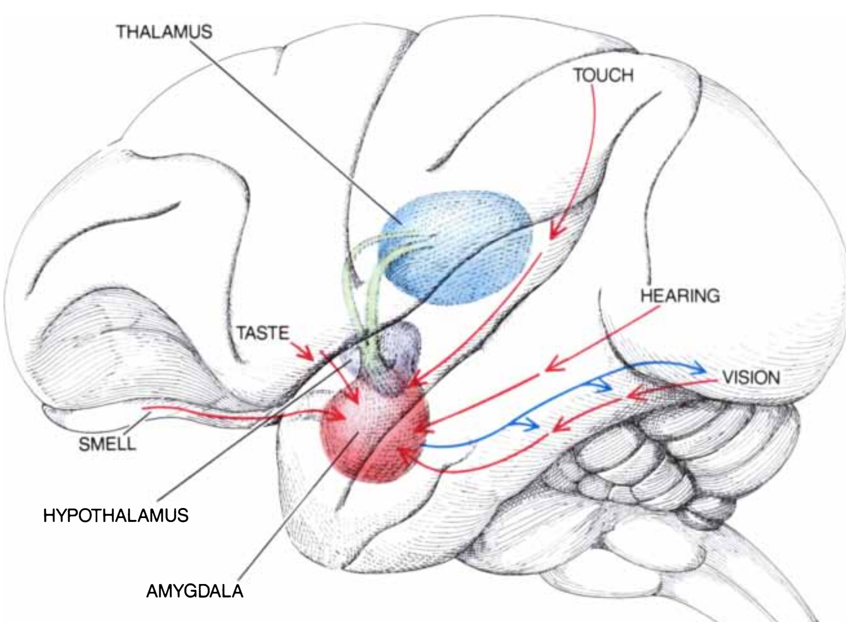
Mixing Memory and Desire

Klüver and Bucy noted another remarkable feature of monkeys lacking temporal lobes, a feature that also was later ascribed to loss of the amygdala. The animals lost their fear of human beings and even their aversion to such normally repugnant sensations as pinching. It was as if a link between familiar stimuli and their emotional associations had been severed. By virtue of its connections with both the sensory areas in the cortex and the triggers of emotional response deep in the brain, the amygdala is well suited to mediating such a link. The possibility that sensory experiences acquire their emotional weight by way of the amygdala gains support from the observation we made early in the course of our inquiry into memory: monkeys without an amygdala are slow in learning to associate an object with a reward. Such animals have trouble remembering the positive associations of a familiar stimulus.

It is possible that the amygdala not only enables sensory events to develop emotional associations but also enables emotions to shape perception and the storage of memories. How does the brain single out significant stimuli from the welter of impressions supplied by the senses? If emotions can affect sensory processing in the cortex, they might provide the needed filter, tending to limit attention—and hence learning—to stimuli with emotional significance. The amygdala, in its capacity as intermediary between the senses and the emotions, is one structure that could underlie such "selective attention."

Circuitry exists that could give the amygdala this gatekeeping function. Several groups of investigators have established that the sensory systems in the cortex not only send fibers to the amygdala but also receive projections from it—projections that, at least in the visual system, are densest in the highest processing stations. In a collaborative study with Candace B. Pert and others in her laboratory at the National Institute of Mental Health my colleagues and I found a clue to the nature of some of the projections. The amygdala is rich in neurons making opiumlike neurotransmitters known as endogenous opiates, which in other parts of the nervous system are believed to regulate the transmission of nerve signals. We found that the sensory processing pathways in the cortex show a gradient of opiate receptors: the cell-surface molecules to which opiates bind in acting on a neuron. The receptors are most abundant in the final stations, where complete sensory impressions take form.

Together, the evidence suggests the possibility that opiate-containing fibers run from the amygdala to the sensory systems, where they may serve a gatekeeping function by releasing opiates in response to emotional states generated in the hypothalamus. In that way the amygdala may enable the emotions to influence what is perceived and learned. The amygdala's reciprocal effect on the cortex may explain why, in both monkeys and humans, emotionally charged events make a disproportionate impression.

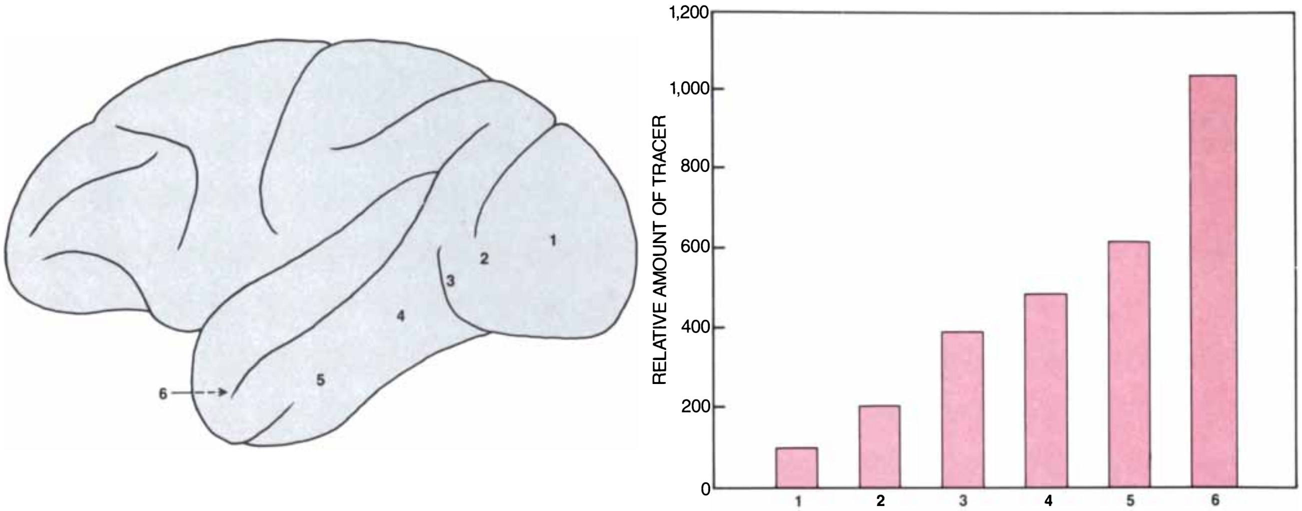


MULTIPLE CONNECTIONS of the amygdala underlie the variety of roles it is thought to serve in memory. Fibers reach the amygdala from the final stations of cortical sensory systems (*red arrows*). Sensory impressions thereby activate one circuit of the memory system, which depends on connections (*green*) between the amygdala and the thalamus (a structure in the diencephalon). Links (*purple*) between the amygdala and the hypothalamus, where emotional responses probably originate, seem to allow an experience to gain emotional associations. Those links may also allow emotions to influence learning, by activating reciprocal connections from the amygdala to sensory paths (*shown in blue, for the visual system only*). The existence of connections running back to the sensory areas from the amygdala may explain why a single stimulus can elicit diverse memories, as when the smell of a familiar food summons memories of its appearance, texture and taste.

Memory and Habit

Having mapped two broad circuits, one rooted in the amygdala and the other in the hippocampus, that are responsible for many kinds of cognitive learning—the capacity to recognize a familiar object, recall its unperceived sensory qualities, remember its former location and attach emotional significance to it—we were left with a puzzle. It is embodied in the fact that people suffering from a memory loss so complete that they cannot recognize another person seen just minutes earlier are still capable of learning. Years ago Milner reported that H.M. learned the skill of mirror drawing (drawing while watching one's hand in a mirror). He mastered it at an almost normal rate, even though after he had done so he could not remember ever having performed the feat.

Monkeys whose limbic structures have been destroyed can also learn. Such animals are helpless when it comes to the delayed nonmatching-to-sample test, in which they must recognize an object seen just once. Yet in our laboratory Barbara L. Malamut has found that if a long series of differ-



EVIDENCE OF A MOLECULAR MECHANISM that may play a role in learning reaches a peak in final stations of the monkey visual pathway. The chart, from a study led by Aryeh Routtenberg of Northwestern University, shows the relative amounts of radioactive phosphorus tracer incorporated into a protein known as *F*1 in tissue from points along the pathway. Routtenberg has proposed that phosphorylation of *F*1 by the enzyme protein kinase *C* underlies changes in the synapses of some neurons after repeated stimulation. Such changes may figure in data storage by the brain. Thus final visual stations may be biochemically suited to memory.

ent object pairs, each containing one baited object, is shown to the same monkeys just once a day, with time they learn to choose the object carrying the reward. What is more, they gain facility at about the same rate as normal monkeys. To a human observer the second task seems, if anything, harder than the first one. How can one reconcile these seemingly contradictory results?

Like many other investigators of memory mechanisms, I have argued for the existence of a second system of learning, one that is independent of the limbic circuits. It is a system for which the critical element is stimulus-response repetition—exactly what is missing in delayed nonmatching-to-sample. In keeping with the evidence from human amnesics, Herbert L. Petri of Towson State University and I have proposed that the second system yields a different kind of learning from the memories stored by way of the limbic circuits.

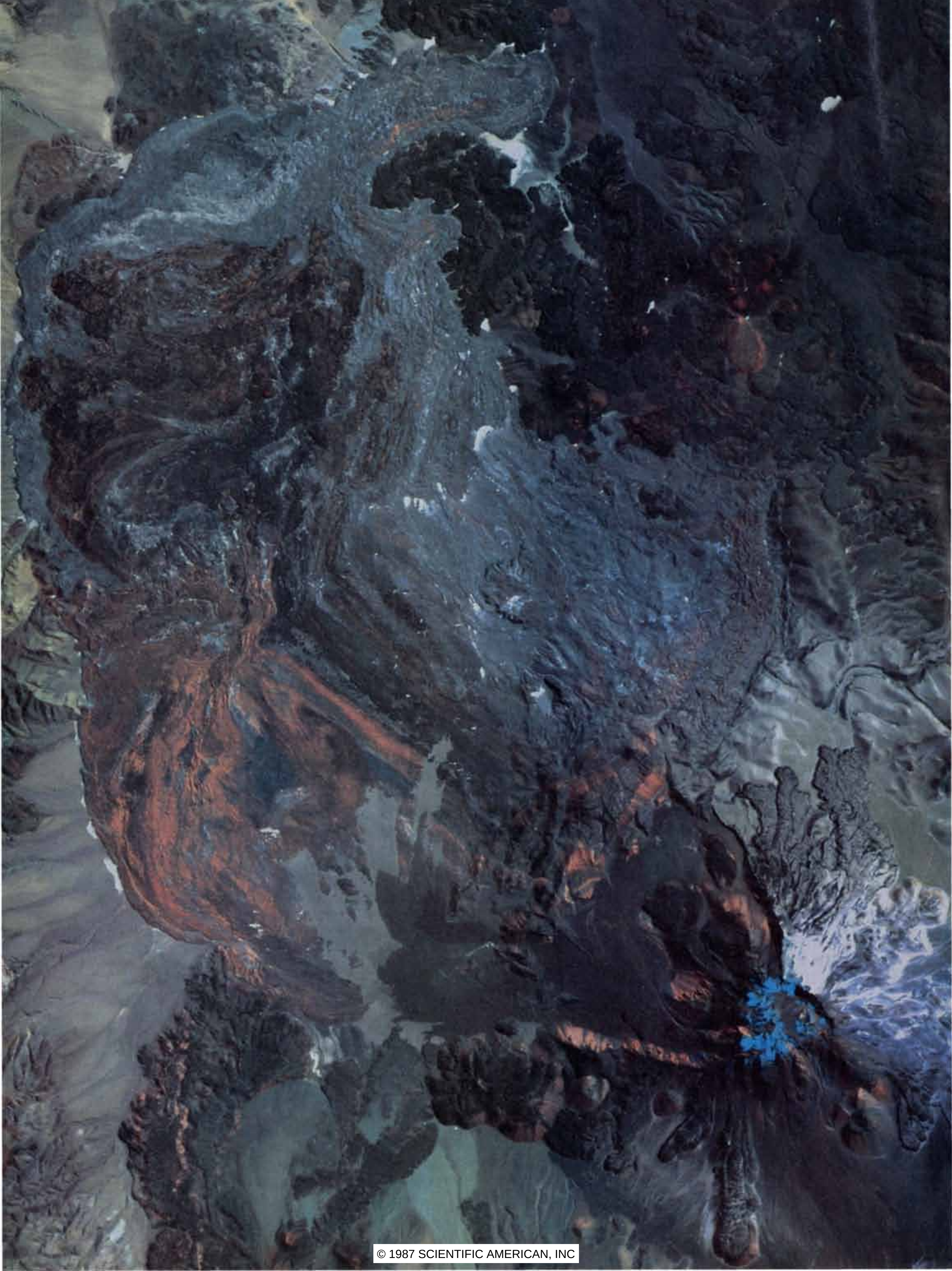
We call this kind of learning "habit." It is noncognitive: it is founded not on knowledge or even on memories (in the sense of independent mental entities) but on automatic connections between a stimulus and a response. In the object-discrimination test the monkey confronts the same pair of stimuli day after day; eventually it develops a habit of picking the object whose selection is always reinforced with a reward. The nonmatching-to-sample task, on the other hand, cannot be satisfied by habit formation. The stimulus to be remembered is shown only once, and the monkey must then respond not to the same stimulus—the one that first carried the reward—but to a new one. It must know, in a cognitive sense, which of the objects is the original one in order to avoid it.

Habits, as we define them, are reminiscent of the automatic stimulus-response bonds that behaviorist psychologists long ago argued are the basis of all learning. The behaviorist point of view excludes such terms as "mind," "knowledge" and even "memory," in its usual sense. It stands in opposition to cognitive psychology, which relies on those very concepts to account for much of behavior. The possibility that learning is built on two quite different systems, one of them a source of noncognitive habits and the other the basis of cognitive memory, offers a way to reconcile the behaviorist and cognitivist schools. If neural mechanisms for both kinds of learning do exist, behavior could be a blend of automatic responses to stimuli and actions guided by knowledge and expectation.

A likely neural substrate for habit formation is the striatum, a complex of structures in the forebrain. The striatum receives projections from many areas of the cortex, including the sensory systems, and sends fibers to the parts of the brain that control movement. Hence it is neuroanatomically suited to providing the relatively direct links between stimulus and action that are implicit in the notion of habit. Indeed, other workers have found that damage to the striatum impairs the ability of monkeys to form habits of the kind that are tested in the object-discrimination task.

Paul D. MacLean of the National Institute of Mental Health has pointed out that the striatum is an evolutionarily ancient part of the brain, older by far than the cortex and the limbic system. One would expect habit formation to be mediated by primitive structures: even simple animals can learn automatic responses to stimuli. Developmentally, habit seems to be primitive as well. Bachevalier has recently found that infant monkeys do about as well as adults on our test of habit formation, and yet they do poorly on the memory test. By adult standards they are amnesic. We are now looking into the possibility that the neural substrate of habit is fully developed in infant monkeys, whereas the memory system is slow to mature. The same developmental difference, if it is present in human beings, could explain why few people remember their infancy.

As for how memory and habit interact in the mature brain, we are only starting to formulate our questions. It seems likely that most kinds of learning draw on both systems, but it is easy to see that cognitive memory and noncognitive habit would often conflict. How does the brain adjudicate between habit formation and cognitive learning? Do elements of the memory system communicate with the striatum and thereby influence habit formation? In surveying the cerebral territory of memory and habit we have only mapped a landscape for future exploration.

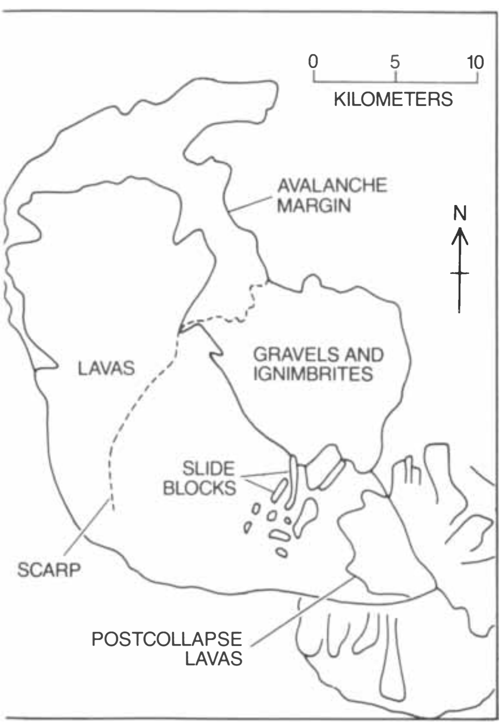


Collapsing Volcanoes

In the life cycle of many volcanoes a catastrophic collapse is a "normal" event. The details of the process are revealed in the deposits left by the devastating avalanches of debris

by Peter Francis and Stephen Self

Volcanoes are unusual mountains. Most peaks are shaped largely by erosion, sculpted by water and ice out of great blocks of tectonically uplifted crust. In contrast, volcanoes are constructional landforms: they build themselves up out of lava and ash. Given that they can actually add to their height, one might ask why they are not consistently the highest mountains on the earth (as indeed they are on Mars and probably on Venus). What factors limit the height of a terrestrial volcano?



SOCOMPA, a volcano in northern Chile that collapsed in a massive avalanche of debris roughly 7,500 years ago, was imaged by the Landsat thematic mapper. The false-color image is a composite of data gathered at several wavelengths. Different debris streams can be distinguished in the avalanche deposit, which covers some 600 square kilometers, on the basis of their spectral characteristics. The gray areas to the north and east are primarily gravels and ignimbrites from the subvolcanic basement; in the reddish areas lavas from the volcano itself overlie the basement material. The blue areas are snow at the top of Socompa, which is 6,051 meters high.

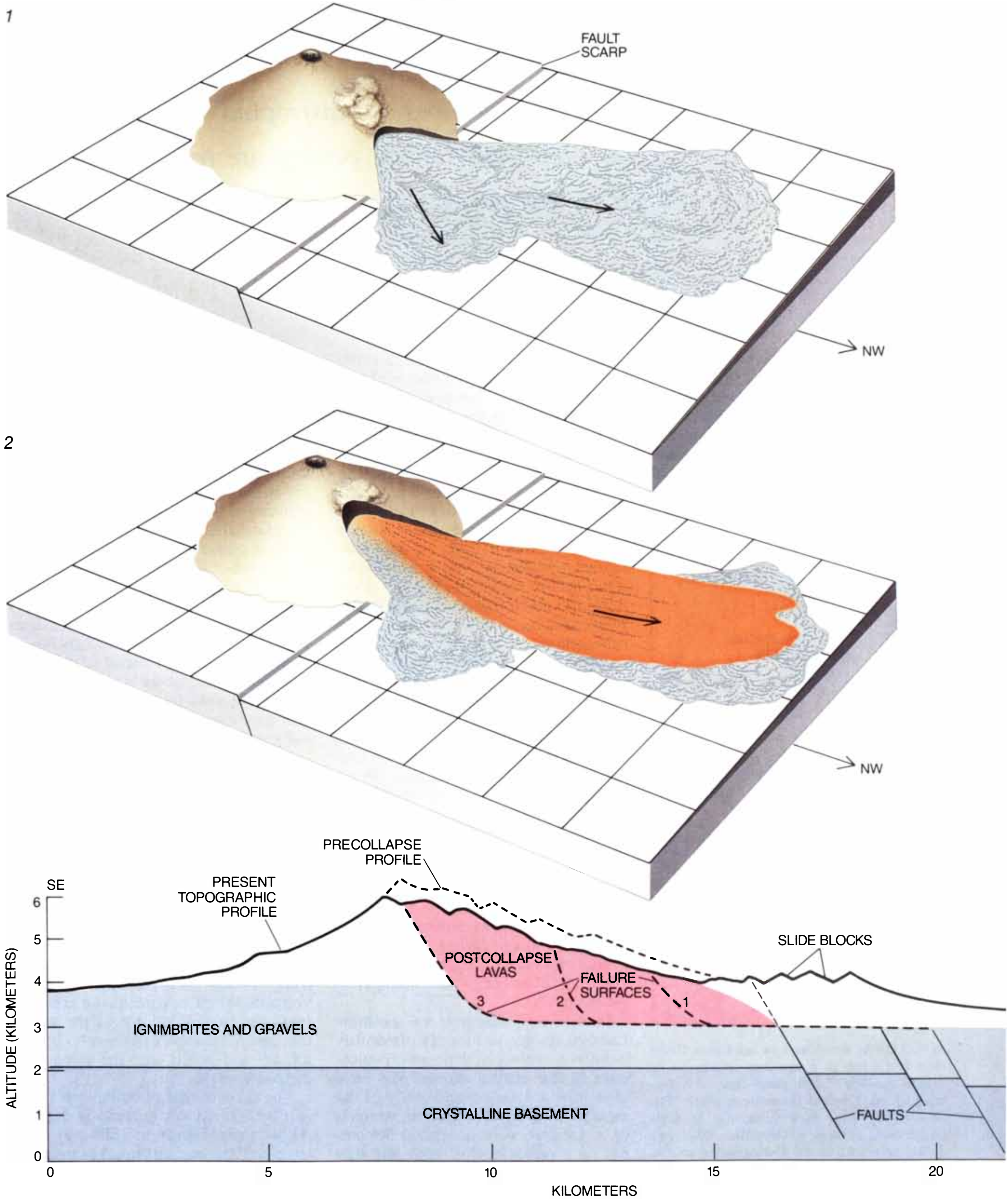
Of course volcanoes too are subject to erosion, as well as to other processes, such as the sagging of the crust under heavy loads, that tend to flatten the earth's topography. Yet in recent years it has become clear that another, more dramatic and distinctive process cuts many volcanoes down to size: wholesale collapse. Although the collapse of a volcanic cone has been observed several times in the past century, it was the catastrophic eruption of Mount St. Helens in May, 1980, that drew widespread attention to the possibility that an entire flank of a volcano could suddenly fail, triggering not only an explosive eruption but also a devastating avalanche of debris. In an effort to understand the collapse phenomenon a number of investigators have turned their attention to the deposits left at Mount St. Helens and at other volcanoes by debris avalanches. More than 100 such deposits have been identified; they range in age from a few years to tens of thousands of years and in size from less than one cubic kilometer to more than 20.

In our own research we examine Landsat images to identify previously unknown avalanche deposits on volcanoes in the central Andes. We have also done a detailed field study of the most striking Andean deposit, which is on a Chilean volcano called Socompa. As a result of this work and that of many other investigators, notably Tadahide Ui of Kobe University, it is becoming clear that the occasional massive collapse of an unstable cone should be seen as a normal event in the life cycle of a volcano. The conclusion holds in particular for the large volcanoes known as stratovolcanoes, of which there are many hundreds in the "Ring of Fire" around the Pacific as well as elsewhere in the world.

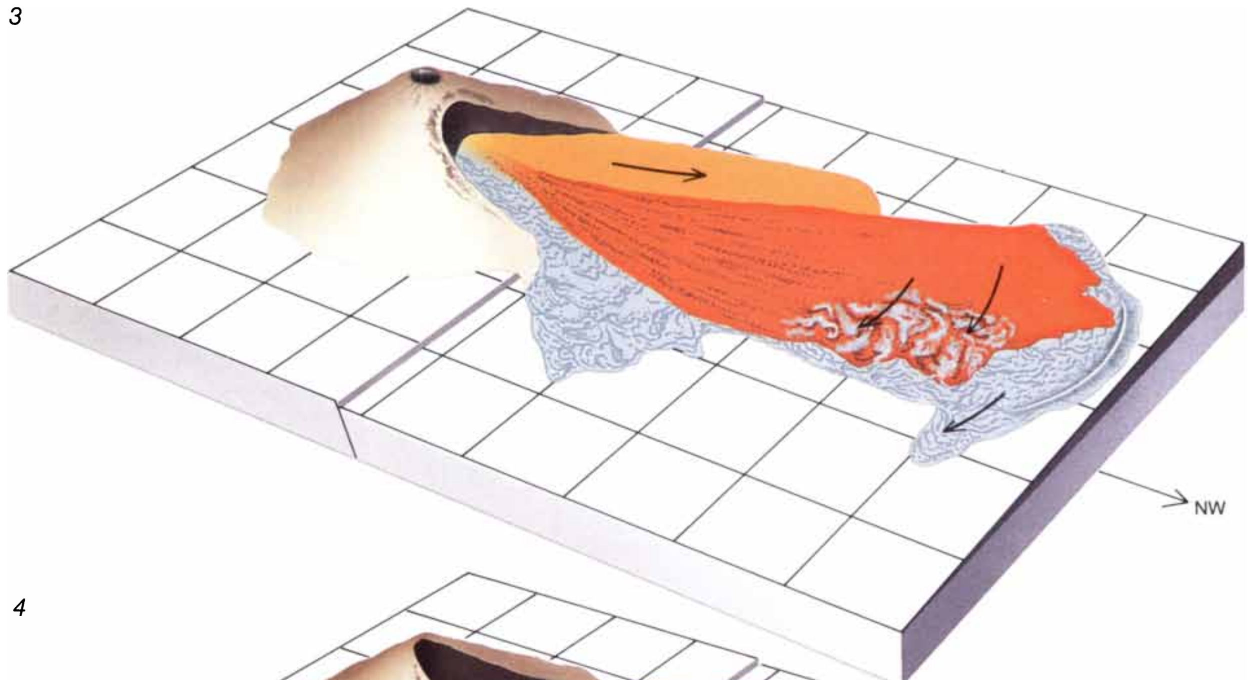
One reason stratovolcanoes are particularly susceptible to collapse has to do with their composition. They are generally associated with subduction zones: boundaries at which two of the plates that make up the earth's outer layer converge and one plunges under the other. (The Ring of Fire is actually a ring of subduction zones.) A subduction-zone volcano tends to emit andesitic or dacitic lavas, which are relatively rich in silica and hence are relatively viscous. Instead of flowing away the lavas pile up around the vent of the volcano, forming a steep cone that may eventually become mechanically unstable.

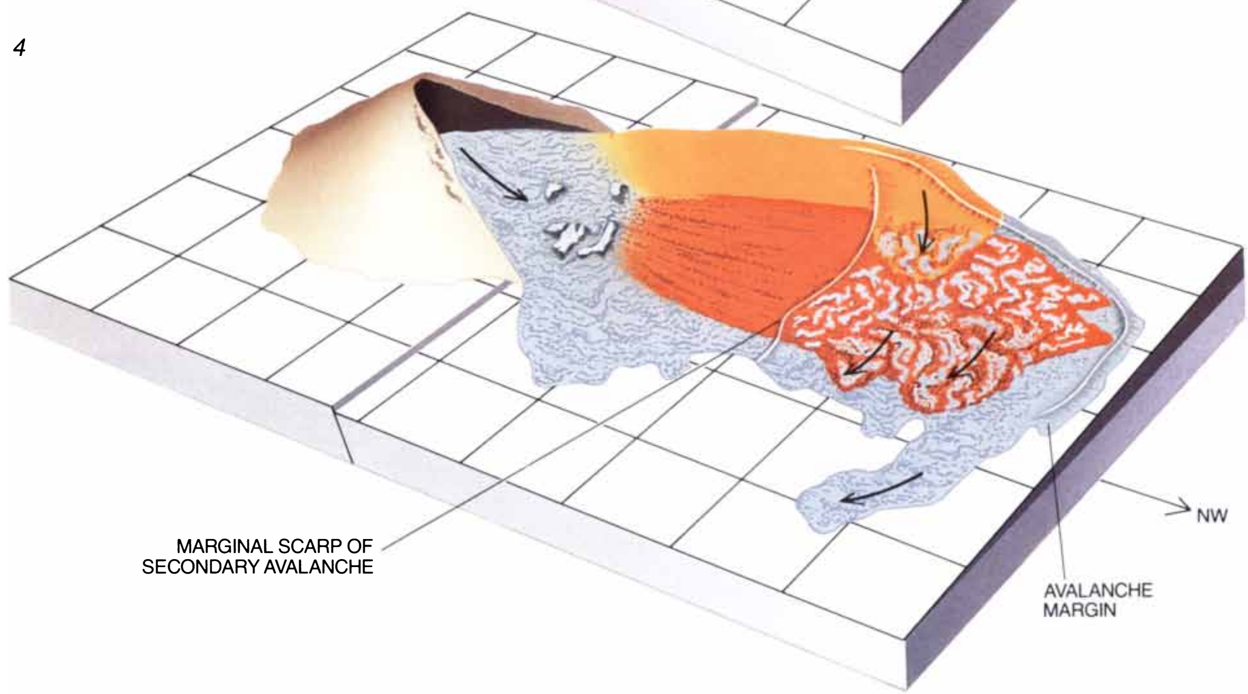
Among stratovolcanoes three kinds of collapse can be distinguished. The first kind is typified by Mount St. Helens, where the collapse set off a violent explosion. At Mount St. Helens a fresh body of magma, or molten rock, had intruded into the volcano. The injection of magma raised a prominent bulge that destabilized the mountain's northern flank; the magma also interacted with groundwater to generate superheated steam, which remained trapped under great pressure inside the volcano. When an earthquake precipitated the partial collapse of the cone, the steam escaped explosively, blasting ash and debris onto the surrounding countryside.

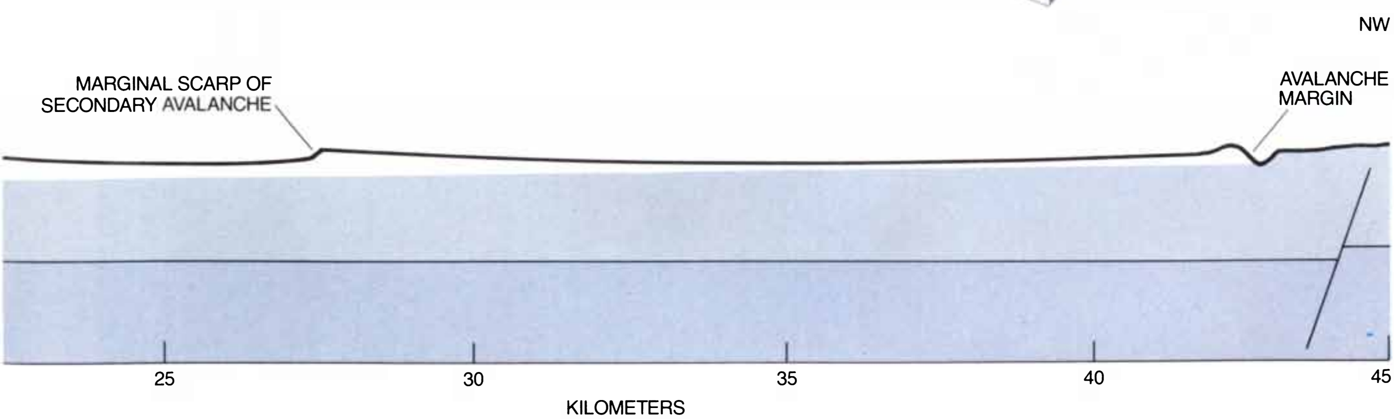
The second kind of collapse is typified by Bandai-san volcano in Japan. Although its failure in 1888 was also triggered by an earthquake and accompanied by explosions, there is no evidence that new magma had intruded into the volcano. The explosions probably resulted from the interaction of groundwater with a hot but no longer molten body of rock. Finally, some collapses have involved no explosion and indeed no volcanic activi-



COLLAPSE OF SOCOMPA proceeded in four stages. The event probably lasted for no more than 10 minutes. Before the collapse the volcano had been destabilized by the intrusion of fresh magma, which may have formed domes on its western flank. The collapse was probably triggered by an earthquake along a fault scarp under the western flank (*1*). The first failure surface was near the base of the mountain (*see cross section*), and so the first wave of avalanche material (*gray*) consisted primarily of gravels and ignimbrites scooped out of the subvolcanic basement. This material traveled to the northwest, straight across the regional slope, at a speed of perhaps 300 kilometers per hour. Even before it reached its northwesternmost point, 35 kilometers from the peak of So-







compa, it was overridden by a new wave of material (*orange*) propelled by the failure of the volcano itself (*2*). The new material contained a high proportion of young lavas. Both avalanche masses came to rest briefly; then much of the material began to slump down the gentle (three-degree) regional slope to the northeast (*3*). At the same time another failure high on the volcano launched a third avalanche mass (*yellow*) to the southwest of the first two masses. Parts of all three continued downslope, overriding and mixing with primary avalanche material and leaving a prominent marginal scarp (*4*). Finally great blocks broke off the walls of the amphitheater carved by the collapse. The blocks slid several kilometers before stopping intact at the mouth of the amphitheater.

ty at all; the volcano simply collapsed in a "cold" avalanche like the avalanches observed on ordinary, nonvolcanic mountains. The most recent example of this process took place at Ontake volcano in Japan in 1984.

Mount St. Helens is the only volcano ever to have been observed closely while it was collapsing. How does one recognize a collapse that took place in prehistoric times? In the few cases where a debris avalanche was the last dramatic event in the history of the volcano, the task is straightforward: the vast, horseshoe-shaped amphitheater left by the avalanche remains prominent for thousands of years. On an active volcano, however, the scar may soon be filled in by fresh effusions of lava. At Mount St. Helens the avalanche scooped some 2.8 cubic kilometers of rock out of the mountainside, creating a giant amphitheater two kilometers wide and 600 meters deep, but today a growing lava dome already occupies about a third of the floor. At Tata Sabaya in Bolivia and Parinacota in Chile, both of which endured massive collapses thousands of years ago, the amphitheaters have since been entirely obliterated by lava flows and domes.

In such cases the evidence of a collapse is found in the deposits laid down by the avalanche of volcanic debris. Identifying the deposits is not always easy: they may be camouflaged by younger soil and vegetation, and they may also be mistaken for the product of more ordinary processes of erosion and deposition. Nevertheless, it is remarkable how often the characteristic structure of an avalanche deposit shows through. Its most distinctive feature is hummocky terrain: thousands of small hills and depressions that may cover tens or even hundreds of square kilometers at the base of a volcano. Hummocky topography has been our chief guide as we have scanned satellite images for evidence of past volcanic collapses in the central Andes.

The most spectacular example we have found is Socompa, a volcano in northern Chile on the Argentine border that collapsed about 7,500 years ago. Whereas the amphitheater at Mount St. Helens occupies about 30 degrees of the volcano's circumference, the avalanche at Socompa carved out a 70-degree wedge, and the volume of material that swept down the mountain's flanks was 10 times as great as at Mount St. Helens. The avalanche deposit is one of the largest in the world, covering an area of some 600 square kilometers. Moreover, the deposit is magnificently preserved. It makes Socompa an exemplary case study of a volcanic collapse.

Currently Socompa is 6,051 meters high, but before the collapse its altitude was probably about 6,300 meters. The mountain is much taller on the western side, where it rises nearly 3,000 meters above the Atacama Desert, than on the eastern side, where the drop from peak to base is about 2,000 meters. The difference in altitudes between the eastern and western base levels is an important clue to what precipitated the collapse. The terrain east of Socompa is higher because it has been uplifted along a series of major extensional faults that run from north to south and that underlie the base of the mountain's western flank—which is precisely where the collapse began. Where there are faults, of course, there are usually earthquakes. It seems quite probable that a powerful earthquake along one of the faults triggered the collapse of Socompa, just as an earthquake triggered the collapse of Mount St. Helens. Traces of the faults are visible today; they cut through recent gravel deposits, which suggests they

"BREADCRUST" BLOCKS of dacite lava in the Socompa avalanche deposit prove that fresh magma was present in the volcano before it collapsed. The glassy exterior and frothy interior of the block in the foreground indicate that it cooled rapidly from a molten state during the avalanche. The hills in the middle distance are the slide blocks, several hundred meters high and as much as two kilometers long, that descended from Socompa in the final stage of the collapse. Socompa's peak is 16 kilometers away.

may still be active 7,500 years after the cataclysm.

For that matter, Socompa must have been shaken by scores of earthquakes during the thousands of years before it fell. Why then did it not collapse earlier? Again the answer seems to be the same as in the case of Mount St. Helens. Field evidence shows that at the time of the collapse fresh magma had intruded into Socompa and may even have been extruding as lava on the high flanks of the cone. The avalanche deposit includes numerous impressive "breadcrust" blocks, so named because their glassy exterior and frothy internal texture make them look like a well-baked, crusty loaf. Some of the blocks are as much as 20 meters in diameter. They could have come only from a body of new magma, which by intruding into the volcano would have tended to destabilize it.

In our view it was this lethal coincidence of an earthquake with a period of magmatic activity that led to the collapse of Socompa. Whether the collapse in turn set off an explosion is uncertain. At Mount St. Helens the distinctive deposit of ash and debris left by the great blast is extremely thin (it ranges in depth from several centimeters to about a meter), and within a few decades it will have been largely stripped away by erosion. It is therefore not surprising that there is no unequivocal evidence of a blast at Socompa. In general, though, the odds are quite high that a large volcanic avalanche will trigger an explosive eruption when new magma is present. The evisceration of the volcano leads to a sudden release of the pressure on the hot interior, and that is a reliable recipe for an explosion.

The avalanche at Socompa took a huge bite out of the mountain. Whereas most of the debris at Mount St. Helens came from the volcanic structure itself, a considerable fraction (probably more than half) of the much larger slide at Socompa was derived not from the volcano proper but from the underlying basement. The basement, which extends downward from an altitude of about 4,000 meters, consists of volcanic rocks from an entirely different geologic era and of some rocks that are not volcanic at all. Hence on the Socompa avalanche deposit one encounters not only the usual lava boulders but also a lot of ignimbrites (rocks formed from the ash of ancient eruptions) and sedimentary gravels. We believe these rocks were disgorged because the failure surface sliced through the western fault zone and cut deep into the basement. The first wave of the avalanche came primarily from the basement rather than from higher up on the mountain.

The dynamics of the avalanche were extraordinary. When the first wave reached the foot of Socompa, traveling toward the northwest, it did not descend the regional slope toward the northeast. Instead it continued on its path, shooting straight across the regional slope until it encountered steeply rising ground about 35 kilometers from the peak of the volcano. (The northwestern limit of the avalanche is marked by a flow front more than 40 meters high.) The fact that the avalanche did not follow a curved trajectory—as a ball would, for instance, if it were rolled slowly across an inclined plane—suggests that it was moving at great speed. There is no way to determine the speed directly, but judging from avalanches that have actually been observed it was probably on the order of 300 kilometers per hour.

The collapse of the subvolcanic basement caused the volcano itself to fail, with the result that the first wave of material was almost immediately overridden by a second wave com-

PARINACOTA, another volcano in northern Chile, collapsed between 15,000 and 17,000 years ago. The photograph shows two characteristics of volcanic avalanche deposits: on the left, at the foot of the volcano, are large slide blocks; in front of them are smaller hummocks. A lake fills the depressions among the hummocks. Judging by the size of the deposit (about 150 square kilometers), the collapse of Parinacota was less substantial than that of Socompa. The amphitheater has been obliterated by lava flows.

ing from higher on the mountain and moving in the same northwesterly direction. Both of these primary avalanche masses came to rest only briefly. Then all the material that was more than about 25 kilometers from the peak, except for a narrow ribbon along the northwestern margin, slid away at a right angle to its initial course, down the regional slope toward the northeast. This secondary flow overrode primary avalanche material, and so its margin facing the mountain is marked by a prominent, northeastward-trending scarp, or cliff. Subsequently parts of both the primary and the secondary flows were overridden by yet another wave of avalanche material from higher on the volcano.

Ultimately some of the material in the Socompa avalanche traveled nearly 40 kilometers before grinding to a halt in the desert. What explains this remarkable reach? One might suppose the answer lies in a massive explosion, if indeed a blast occurred at Socompa. It turns out, however, that even large explosions do not contribute much, if anything, to the energy of a volcanic avalanche. In this respect volcanic avalanches are like ordinary ones: their energy comes primarily from gravity. In other words, avalanches that start higher have the potential to travel farther. At Socompa the steep western slope clearly helps to explain why the avalanche flowed far into the Atacama Desert.

But the steep slope is not explanation enough. The Socompa avalanche flowed much farther than one would expect from the topography of the region. Much of the material in the avalanche came from the subvolcanic basement, which means that it started at an altitude of 4,000 meters or less. The northwestern margin of the deposit lies about 30 kilometers from the base of the mountain, at an altitude of 3,200 meters. Dividing the vertical drop by the horizontal distance traveled, one arrives at a ratio of less than .03. (The calculation for the secondary, northeastward flow yields a strikingly similar result.) The ratio of vertical to horizontal distance is a good indicator of the mobility of an avalanche; a typical value would be about .1. The Socompa avalanche seems to have been unusually mobile.

The most plausible explanation is that the earthquake that triggered the collapse of the volcano also mobilized the avalanche. During a major earthquake (one with a Richter magnitude of seven or eight) severe ground shaking may continue for several minutes; moreover, the shaking can be intensified and prolonged when the seismic waves are bounced back and forth in a basinlike configuration of subsurface rocks, as is present at Socompa. The Socompa avalanche would have needed only a few minutes to reach its most distant point. The extra seismic energy would have kept the avalanche moving by helping it to overcome the loss of energy to friction. In particular, the ground shaking would have dislodged the primary avalanche masses after they had come briefly to a halt, and it would thereby have started the secondary flow of debris down the gentle regional slope.

The picture we have presented—that of a catastrophic avalanche, stirred by violent tremors as it sweeps down a mountainside at tremendous speed—might seem to suggest that the flow of

MOUNT ST. HELENS AMPHITHEATER is 600 meters deep and two kilometers wide. The photograph was made in September, 1984, by Lyn Topinka of the U.S. Geological Survey. The lava dome growing in the amphitheater now covers a third of the floor.

SCIENTIFIC AMERICAN

In Other Languages

LE SCIENZE
L. 3,500/copy L. 35,000/year L. 45,000/(abroad)
Editorial, subscription correspondence:
Le Scienze S.p.A., Via G. De Alessandri, 11
20144 Milano, Italy
Advertising correspondence:
Publietas, S.p.A., Via Cino de Duca, 5,
20122 Milano, Italy

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

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

POUR LA SCIENCE
27FF/copy 265FF/year 315FF/year (abroad)
Editorial, subscription, advertising correspondence:
Pour la Science S.A.R.L.,
8, rue Férou,
75006 Paris, France

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

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

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

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

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

Advertising correspondence all editions:
SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue
New York, NY 10017
Telephone: (212) 754-0550 Telex: 236115

The advertisers listed below are making additional information available, free. Circle the corresponding number on the adjoining card, detach and mail.

1. Alcoa
2. American Mensa, Ltd.
3. AT&T Technologies
4. Audio Forum - Cassette /Book Language Courses
5. DuPont Company
6. Gould Inc./NPL Mini Supercomputers
7. Interplak/Home plaque removal instrument. Cleans teeth virtually plaque-free.
8. McDonnell Douglas
9. 3M/Data Cartridge
10. New Jersey Department of Commerce & Economic Development
11. NordicTrack/Home cardiovascular exerciser. Send for free brochure.
12. Sci-Mate Manager
13. U.S. Environmental Protection Agency

'SCIENCE OF MAKING MONEY'

14. Citibank Precious Metal Center
15. Dow Jones News Retrieval
16. IBIS
17. Salomon Brothers
18. UNISYS
19. Vanguard Quantitative Portfolio

European vacations so great, we wrote a book about them.

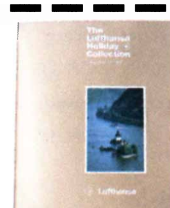
The Lufthansa Holiday Collection® has over 80 pages of European vacations that contain all the drama and intrigue of a good novel.

You'll become easily engrossed as you read about vacations that take you to the heart of Europe—where you can visit centuries-old castles for dinner, tour towns that recreate the Middle Ages and stop at fascinating boutiques, museums and palaces.

Read about all our vacations from cover to cover. For a free copy of The Lufthansa Holiday Collection, call 1-800-645-3880, send in our coupon or see your travel agent.

You won't just become well-read, you'll become well-traveled.

The Lufthansa Holiday Collection.®



Send to: Lufthansa German Airlines
Dept. UX 12, 11847 Teale St.
Culver City, CA 90230
or call toll-free 1 800-645-3880

Name (please print)

Address

City State Zip

Travel Agent

SA 6/87

Lufthansa

BANDAI-SAN VOLCANO in Japan collapsed on July 15, 1888. This contemporary lithograph based on photographs and sketches made three weeks after the collapse shows with remarkable accuracy the debris-filled amphitheater (with steaming vents) and the hummocky topography of the avalanche deposit. Unlike Mount St. Helens, Bandai-san had not been destabilized by new magma.

debris resulting from a volcanic collapse is highly disorganized and turbulent. Such is not the case. The flow is indeed complex, particularly once it has reached top speed, but it is not chaotic or turbulent. The hallmark of turbulent flow, which is observed chiefly in fluids of low viscosity such as air or water, is that individual particles in the flowing mass are constantly in motion with respect to one another. In contrast, the flow in a volcanic avalanche tends to be laminar, that is, it proceeds in smooth layers that retain their relative positions. Only the bottom layer, the one in contact with the ground, is heavily sheared.

Indeed, the motion of the avalanche on the first part of its path, on the slope of the volcano, can be more accurately described as a slide rather than a flow. Enormous blocks of rock become detached from the mountain and accelerate downhill. Although some backward rotation and general jostling take place, the blocks remain intact. During this phase, if one had the nerve, it would probably be possible to ride safely on top of a sliding block. At Socompa some of the blocks—the ones that broke off from the summit region near the end of the avalanche—are preserved at the mouth of the amphitheater, several kilometers from where they started. The largest of them is approximately two kilometers long and half a kilometer high.

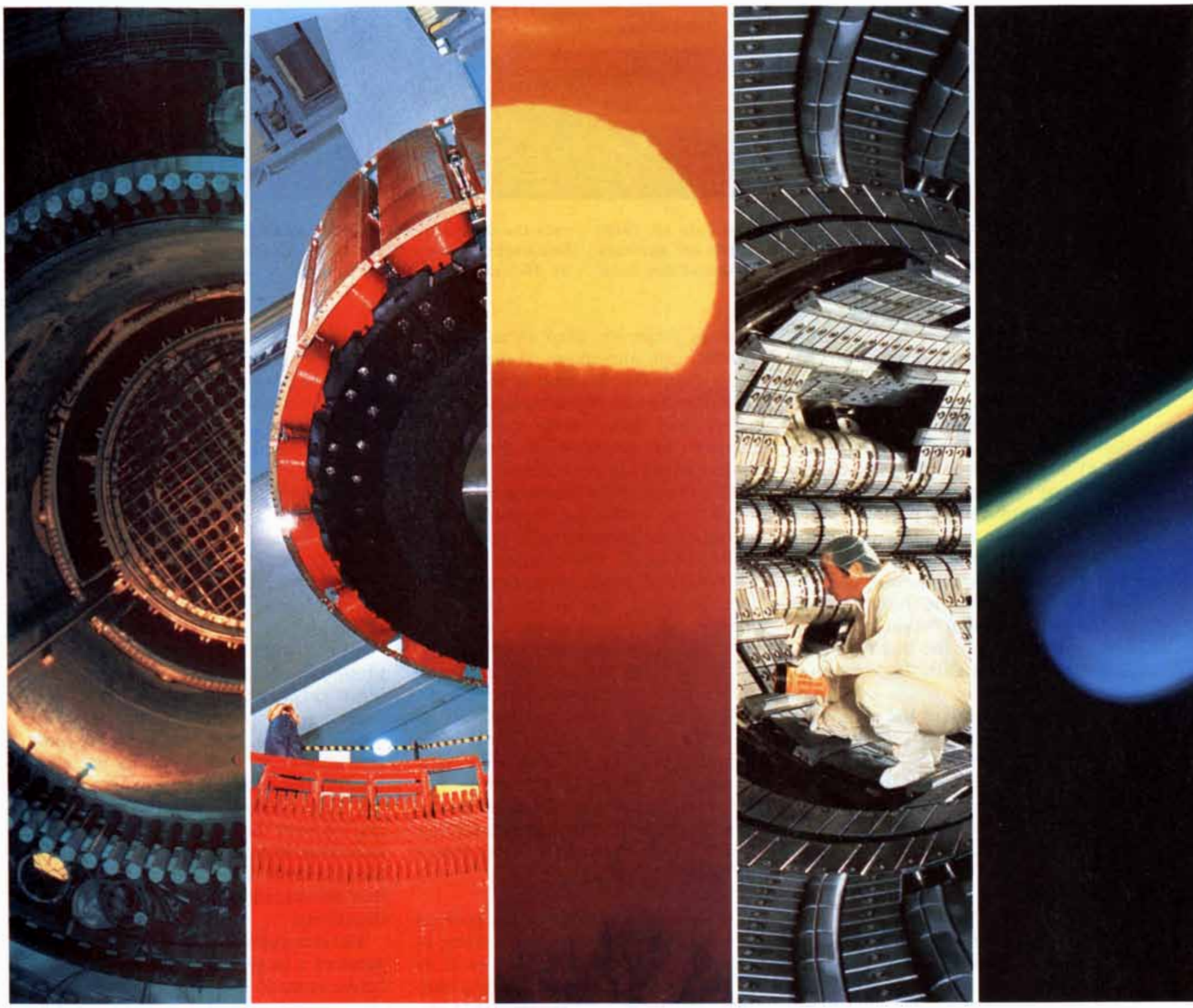
Large blocks undoubtedly break off a volcano throughout an avalanche, but the earlier ones are not stopped at the mouth of the amphitheater and do not survive intact. By the time a block reaches the foot of a volcano, it is moving at a speed of between 100 and 200 kilometers per hour. The shearing stress at such speeds is enough to shatter the block into smaller fragments, most of them less than a meter in diameter. The movement of these rocks away from the volcano is more of a flow than a slide. Nevertheless, it is not a turbulent flow; subtle stratigraphic relations are preserved in the deposit as they existed in the volcano.

This observation has been made at a number of avalanche deposits, and it is true on an exceptionally large scale at Socompa. In the original volcano light-colored dacite lava rested above dark andesite, and in the avalanche deposit the remains of the dacite can be found above the shattered remains of the andesite. Under both, and most visible at the farthest margin of the deposit, are light-toned gravels and ignimbrites from the subvolcanic basement. All three layers have been heavily sheared; whereas originally they were tens or even hundreds of meters thick, in the avalanche deposit they are reduced to thin screens no more than a few meters thick.

Nothing can survive in the path of a volcanic avalanche traveling at high speed and capable, because of its enormous momentum, of rushing hundreds of meters uphill. Furthermore, the destruction may extend to areas far beyond the one covered by the avalanche itself. By disrupting the regional drainage system and causing lakes and rivers to spill over their banks, an avalanche may generate voluminous mud flows. Worse, if the volcano is on the shore of an ocean or a large lake, the sudden displacement of water by the avalanche may trigger a devastating tsunami, or tidal wave. The collapse of the Rakata cone on Krakatau in 1883 is an infamous example. As debris from the cone plunged into the sea at the height of the magmatic eruption, it set off the "great wave" that ravaged several hundred kilometers of coastline in the Sunda Strait (between Java and Sumatra), washing away entire towns and killing some 30,000 people.

Most volcanic collapses are probably preceded by magmatic activity and accompanied by an explosive eruption. In such cases the shape of the volcano may offer ample warning of a possible cataclysm. Before the eruption of Mount St. Helens, for instance, volcanologists who were monitoring the growth of the bulge on its northern flank were acutely aware that the mountain might fail (although with hindsight it is fair to say that they underestimated the scale of the collapse). When a volcano shows the potential for failure, it is possible to prepare hazard maps indicating the areas that are most likely to be affected by the explosion or the avalanche. For understandable reasons, workers who study these hazards tend to focus their attention on volcanoes with a track record of activity.

Yet it is quite possible, given the impetus of a large earthquake, for an inactive volcano to collapse without obvious warning. Such a cold avalanche may be as destructive as a major eruption. In 1792, for example, the eastern flank of Unzen volcano in southern Japan collapsed into the Ariake Bay, and the ensuing tsunami killed more than 14,500 people. Until the shape of a conical volcano has been degraded and stabilized by erosion, it must be regarded as a candidate for collapse, whether or not it is active. That observation adds a dimension to the threat volcanoes pose to the world's inhabitants, particularly to those who live on the rim of the Pacific Ocean, where both volcanoes and earthquakes are commonplace.



Hitachi's wide-ranging technologies in energy (from left to right): nuclear power reactor, generator-motor, solar energy development, nuclear fusion plasma testing device, and laser-test of liquefied petroleum gas combustion.

Diving Adaptations of the Weddell Seal

Collapsible lungs and a spleen that functions as a scuba tank are apparently among the features that enable the seal to swim deeper, and to hold its breath longer, than most other mammals

by Warren M. Zapol

A person who can swim unaided to a depth of 20 meters and stay submerged for three minutes is considered an expert diver. Yet such an accomplishment pales when compared with that of another mammal, one able to plunge more than 500 meters and remain underwater for more than 70 minutes. This diving virtuoso is the Weddell seal (*Leptonychotes weddelli*), a member of the Phocidae family of true, or earless, seals.

The animal, which flourishes on the shores and coastal ice of Antarctica, plows deep into the cold sea not to set endurance records but in search of food. A quarter of a mile from land, within 50 feet of the 250- to 600-meter-deep sea floor, lives its staple diet: the large Antarctic cod *Dissostichus mawsoni*.

Weddell seals readily withstand water temperatures that fall to −1.9 degrees Celsius by virtue of their large size (adults weigh from 350 to 450 kilograms) and a thick layer of insulating blubber. Diving, which forces the animals to cope with a lack of air and with intense undersea pressure, constitutes a more complex challenge. Indeed, unraveling the adaptations to this challenge has required decades of laboratory research by many investigators and, more recently, a spate of field studies. The field studies suggest that certain long-held beliefs based on laboratory studies may need to be modified. Forcing a seal confined in a laboratory to put its face underwater does not necessarily evoke the same response as a dive undertaken freely in the sea.

The specific problems posed by diving are considerable. Above all, the seal must provide its tissues with oxygen. At the same time it must limit the buildup in the blood of carbon dioxide, a by-product of the oxidation of glucose for energy. This gas is generated by the tissues and then carried in the blood to the lungs for removal. When an animal is submerged, the gas can accumulate in the blood, upsetting the fluid's delicate *p*H balance.

The animal also has to avoid the many ills that extreme pressure can cause. For every 10 meters of depth, an animal or a person is subject to an additional "atmosphere" of external pressure—that is, to the push exerted at sea level by a 14.7-pound force on one square inch of area, or the pressure exerted by a 760-millimeter column of mercury. One potential effect of underwater pressure is an increase in the excitability of nerve cells, which can result in convulsions. Pressure also squeezes air pockets, such as the air sinuses in the human head. The squeezing can cause pain, and if the body cannot deliver enough air to the pockets to equalize the external pressure, blood vessels may expand into the air spaces and burst.

Pressure also compresses gases, posing a danger when it affects the nitrogen in the alveoli, the tiny gas sacs in the lungs. (Body fluids and fluid-filled organs are compressed only minimally underwater.) Nitrogen gas constitutes some 78 percent of the air. Normally it passes harmlessly into the circulation, but when the air in the lungs is put under great pressure, as it is during descent, excess nitrogen dissolves in the blood and tissues; it may then lead to narcosis, a disorder that divers call rapture of the deep. Narcosis is identified by such symptoms as intoxication, loss of coordination and vision, drowsiness and unconsciousness. During ascent, a too rapid trip to the surface can cause the nitrogen tension in the blood and tissues to be greater than the external pressure on the body. Then dissolved nitrogen may come out of solution and bubble ("the bends"). In addition to producing pain in the joints and elsewhere, the bubbles may block vessels in the brain and spinal cord, leading to paralysis and even death.

Laboratory studies, in spite of their limitations, have revealed many of the strategies by which the diving seal appears to ensure an adequate supply of oxygen and avoid the disorders described above. I shall therefore review those strategies and then indicate instances in which field studies contradict or clarify earlier results.

One major and still undisputed laboratory finding is that the seal stores an abundance of oxygen—almost twice as much per kilogram of body weight as a human being does. It also concentrates the oxygen where it is most needed during a dive: in the blood and, to a lesser extent, in the muscles. People are particularly dependent on the lungs for oxygen, keeping 36 percent of their total supply in the lungs and 51 percent in the blood, but the seal stows only 5 percent in the lungs and a full 70 percent in the blood. Similarly, a person stores just 13 percent of its oxygen in the muscles, but the Weddell seal keeps about 25 percent there, bound to the oxygen-carrying pigment myoglobin.

Vast amounts of oxygen can be maintained in the seal's blood in part because the volume is enormous. In 1969 Claude J. M. Lenfant, then at the University of Washington, discovered that in contrast to the blood of human beings, which typically accounts for 7 percent of the body weight, the blood of the Weddell seal accounts for 14 percent of the animal's weight. (This comparison actually underestimates the amount of blood that is available to working tissues, because blubber, which constitutes about a third of the animal's mass, receives little blood.) Moreover, the seal's blood has great quantities of hemoglobin, the oxygen-carrying pigment of red blood cells. When my group at the Massachusetts General Hospital drew blood from seals in the laboratory, we found that red cells accounted for some 60 percent of the volume of each drop; in human beings these cells occupy only from 35 to 45 percent of the volume.

Although the Weddell seal's oxygen supply is impressive, it is not infinite. Like other diving animals, the seal has therefore devised ways to conserve its fuel. When any mammal puts its face in the water, neural impulses trigger the brain to induce the so-called diving reflex: as the animal stops breathing, bradycardia (a slowing of the heart rate) ensues and certain arteries become constricted, limiting the blood that flows to the organs they feed.

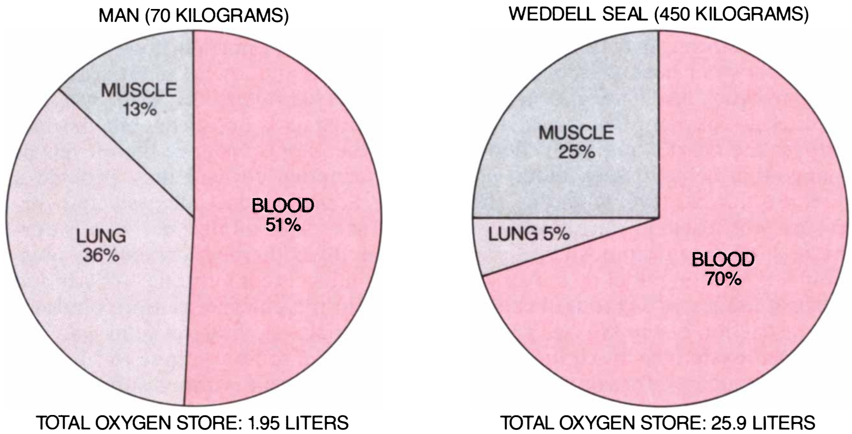
The rapid onset of bradycardia at the start of a dive has been recognized in animals for more than 100 years. It happens in human beings but appears to be most profound in species that dive habitually, such as seals and whales. A slowed heart rate is beneficial underwater because it enables the heart to work less hard and hence to require less oxygen. Bradycardia also reduces the heart's output of blood, a change that helps to keep blood pressure at a normal level when the arteries are constricted. Furthermore, as the flow of blood diminishes, the metabolism slows, reducing the oxygen needs of tissues throughout the body.

Constriction of the arteries presumably ensures that the maximum supply of blood, and therefore of oxygen, will be available to the tissues that are most crucial to survival. My colleagues and I recently measured blood flow to various tissues in the seal during a laboratory dive. Consistent with earlier findings, we found that the seal continued to supply blood at a normal rate to the retina, brain and spinal cord, all of which are vital to navigation and motor control. (As would be expected, the heart received blood but the amount was reduced to match the organ's reduced workload.)

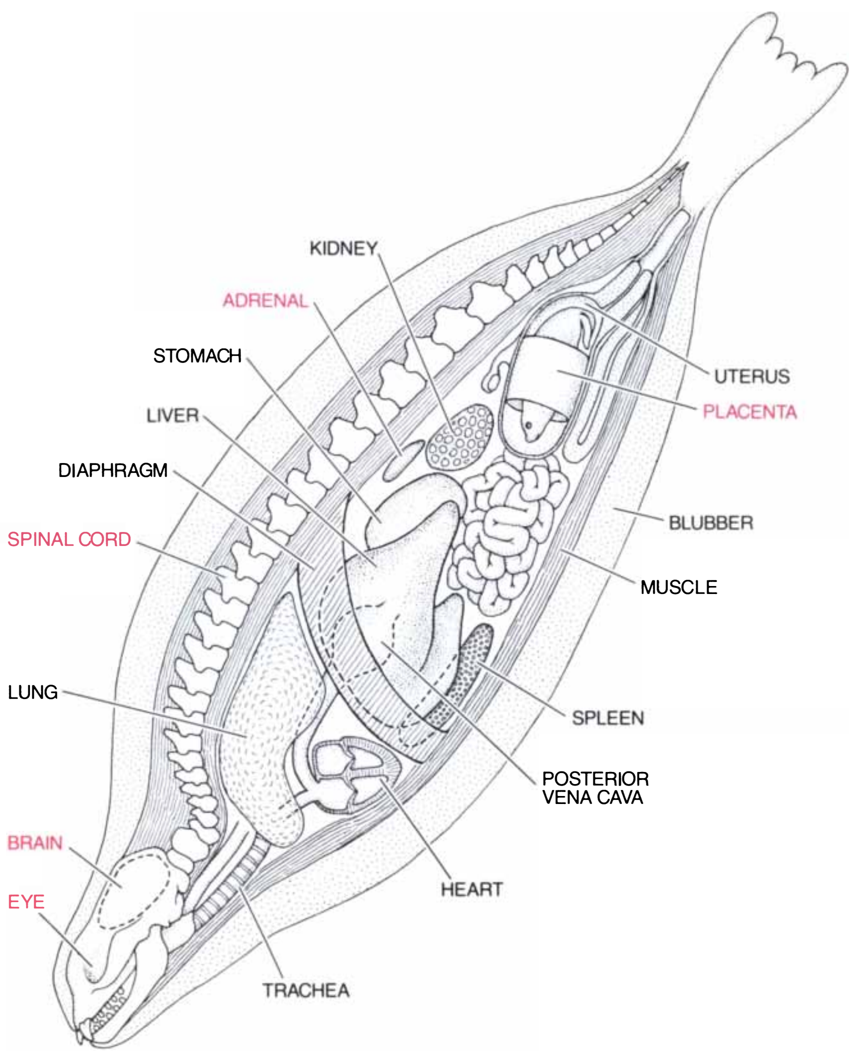
Two other tissues also continued to benefit from a normal flow: the adrenal glands and, in pregnant seals, the placenta. Just why the body perfuses the adrenals is not clear, but the fact that the glands produce high levels of the hormone cortisol may provide a clue. Some evidence suggests that cortisol serves to stabilize nerve cells during a dive, thereby preventing pressure-induced convulsions. Why blood flows to the placenta is more obvious. This organ is vital to fetal gas exchange and must continue to function if the submarine within a submarine is to survive.

Our studies of oxygen distribution also confirmed that the seal essentially shuts off the flow of blood to most other organ systems and tissues during laboratory dives. When this flow ceases, many of the affected tissues (such as the kidneys) stop functioning until the animal comes up for air. Certain other tissues apparently switch to anaerobic, or oxygen-independent, metabolism if they have cru-

WEDDELL SEAL returns to a manmade ice hole for air after a dive in the Antarctic Ocean. Most dives are fishing trips that last for about 17 minutes and take the animals as deep as 500 meters or more. Other dives may last for more than an hour. To survive even short forays the animal must supply oxygen to its tissues and avoid such pressure-associated ills as nitrogen narcosis and decompression sickness ("the bends"). Randall W. Davis of the Sea World Research Institute in San Diego made the photograph.



DISTRIBUTION of oxygen differs markedly in a man and a Weddell seal. The seal stores about twice as much oxygen per kilogram of body weight as the person. The animal is also less reliant on its lungs, keeping more of its oxygen store in its blood and muscles.



TISSUES that continue to receive normal amounts of oxygen-laden blood (*colored labels*) during laboratory dives include the retina of the eye, the brain and the spinal cord, all of which are crucial to navigation or motor control. The adrenal glands, which produce a hormone that may protect the brain from pressure, also receive a normal flow, as does the placenta in pregnant seals. Not knowing how long they will be underwater, the seals apparently prepare for the worst by dramatically reducing blood flow to many other tissues, including their extensive muscular system. When oxygen stores have been depleted, the muscles obtain energy by initiating anaerobic, or oxygen-independent, metabolism.

cial tasks to perform. The telling by-product of anaerobic metabolism is lactic acid; when seals surface after forced dives, the levels of lactic acid in the blood soar above the resting state.

Although initiating anaerobic metabolism can be important when oxygen is lacking, it can also be extremely dangerous. High levels of lactic acid lower the *p*H of the blood and can lead to acidosis, which may cause cramping, a weakening of the heart's ability to contract and even death. Laboratory studies by P. F. Scholander of the Scripps Institution of Oceanography, who studied the diving reflex in the 1930's, suggested that the seal avoids acidosis by confining anaerobic metabolism to the skeletal muscles and other tissues that are isolated from the blood supply during laboratory dives. With the blood flow shut off, these tissues cannot release lactic acid into the blood until the animal surfaces. At that time the liver, lungs and other organs can clear out the by-product.

Laboratory work has also attempted to explain how the Weddell seal handles external pressure. In addition to raising the possibility that elevated cortisol levels may prevent convulsions, the studies have shown that the seal lacks the potentially troublesome air sinuses of other mammals. The seal likewise has ways of avoiding nitrogen narcosis and the bends. Its lungs are small for its weight, and so they have a reduced capacity for storing nitrogen that might diffuse into the blood in the course of a dive. Moreover, the animal exhales before submerging. The obvious effect is to reduce the buoyancy that impedes descent, but exhalation has the added benefit of reducing gas volume in the lungs still further.

During a dive, seawater pressure on the animal's collapsible rib cage undoubtedly squeezes most of the remaining nitrogen out of the alveoli and into the bronchial air-duct system. According to anatomical studies done by Gerald L. Kooyman and his co-workers at the Scripps Institution, the seal's bronchi and bronchioles are supported by rings of cartilage that enable the airways to serve as an armored gas-storage reservoir. Because these passages, unlike the alveoli, have no direct contact with the blood, they do not introduce nitrogen into the circulation. (Some oxygen is certainly sequestered in the seal's airways as well, but not much; only 21 percent of inhaled air is oxygen.) In contrast, the bronchi and bronchioles of human beings would close down under intense pressure and so could not store excess nitrogen.

Several years ago Kooyman also determined, on the basis of forced dives in a compression chamber, that the seal's lungs collapse when the animal reaches a depth ranging between 50 and 70 meters. Collapsed lungs would halt the flow of nitrogen into the blood and hence limit the total amount of nitrogen that accumulates there.

More recently Kooyman and his coworkers have carried out some field studies that have caused many of us to wonder whether the seal in the ocean responds to a dive the way it does in the laboratory. When Kooyman's group attempted to study seals engaged in voluntary dives in the ocean, their data suggested that the animals may not always exhibit a pure diving reflex. Those early studies were not definitive, however, because the available equipment could not monitor complete dives.

After Roger D. Hill of Massachusetts General developed software and constructed a battery-operated, eight-bit computer that would make it possible to evaluate the seal's physiological and metabolic responses throughout free dives at sea, a group of us from laboratories around the world converged on the National Science Foundation's research station on the shore of McMurdo Sound in Antarctica. With any luck, the portable computer, which weighed less than four pounds and had 64 kilobytes of random-access memory, would enable us to clarify the extent to which the Weddell seal exhibits the diving reflex in its natural habitat. It would also add insight into how the seal deals with pressure.

Hill's diving computer, which was encapsulated to withstand 500-meter depths, did everything but steer. It recorded heart rate and depth at predetermined intervals for several days. It also controlled an electric pump that took up to seven arterial-blood samples at specified times (such as 10 minutes into the dive) and depths. After collecting the samples the computer pumped the blood into a bag or syringes tethered to a six-foot fiber-optic line, which itself had additional functions when the seal surfaced.

We gathered seals from nearby colonies and sledged them to the study site, a hole three feet in diameter drilled through ice six feet thick. There we anesthetized a subject with harmless techniques devised by Robert C. Schneider of Massachusetts General, inserted the necessary catheters and attached the computer, which was fastened to a rubber sheet glued to the seal's dorsal fur. (When the animals molted later in the summer, they readily shed the rubber appendage.)

Once rigged with our computer and recovered from general anesthesia, the seal was free to enter the hole and swim off. We were confident it would return with blood samples and data because we adopted a rather reliable tactic devised by Kooyman. Knowing that Weddell seals can swim only a few kilometers underwater, he studied the animals at isolated holes drilled in broad ice sheets; the seals had to return to the site of origin in order to breathe.

After a subject went on its way, we placed a fishing hut (with a large circle cut out of the floor) over the hole, providing shelter both for us and for a computer that would later retrieve data from the diving computer. When the seal returned, we quickly connected its fiber-optic line to the stationary computer. Within 10 seconds the larger machine collected data stored in the diving computer and, when appropriate, gave it new instructions.

We found, as Kooyman had earlier, that some 95 percent of the seal's voluntary dives last for less than 20 minutes. These tend to be feeding dives in which the animals head directly for their prey and then return. The seals embark on the 5 percent of dives that last longer than 20 or 30 minutes when they explore distant routes or must escape from predators.

ATTACHING a combination computer and blood sampler to the back of a seal enabled Roger D. Hill of the Massachusetts General Hospital (*right*), the author and their colleagues to measure the physiological responses of seals throughout free dives at sea. Journeys shorter than 17 minutes evoke a less dramatic response than dives in the laboratory.

Studying the all-important distribution of oxygen to tissues, Peter W. Hochachka of the University of British Columbia showed that the seals do not release lactic acid into the circulation during or after sea journeys that last for up to 20 minutes. This indicated that during short natural dives—that is, the majority of the seals' journeys—the muscles do not resort to the anaerobic metabolism observed in laboratory dives and must receive some blood. (The muscles would probably account for most of the lactic acid in the blood because they are abundant and also do work when the animal dives.) With little or no lactic acid to break down after a dive, the seal often resumes fishing within minutes after taking a few breaths at the surface.

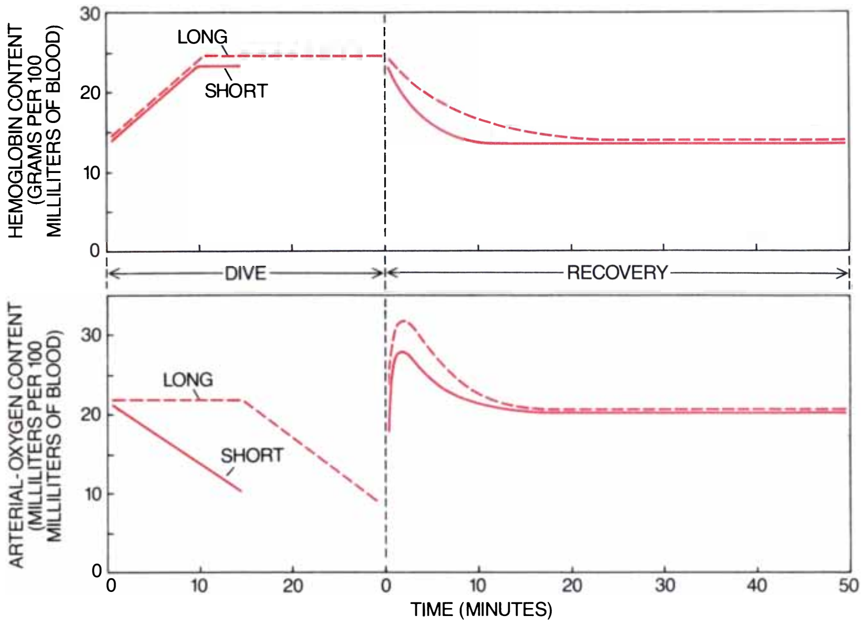
We wondered how Weddell seals supply oxygen to the muscles without depriving the brain and other vital tissues of their rightful supply. No one yet knows the answer, but Hill did find a hint. He noted that the heart rate slows at the start of every dive but does not remain at a constant level throughout the shorter excursions; instead it quickens and slows in accordance with the seal's swimming speed, never outpacing the resting rate. When the heartbeat quickens, cardiac output must also increase. The extra blood has to go somewhere, and it may well be channeled to the skeletal muscle. If this is the case, it implies that the total constriction of blood flow to muscles, assumed for so long to characterize every dive, does not in fact take place in the course of most natural dives.

In contrast to the feeding dives, the seal's occasional long excursions do evoke the classic diving response seen in the laboratory. The long forays are characterized by profound bradycardia with little variability of heart rate. After (but not during) such dives the Weddell seal releases lactic acid into its blood, indicating that the animal shuts off the blood flow to its muscles and meticulously conserves oxygen while diving. Having switched to anaerobic metabolism in the muscles, the seal can stay underwater for an hour or more. It pays a price, though: when it finally surfaces, it does not dive again until it has cleared away the lactic acid released by the muscles, a process that can take up to an hour. Why do even short laboratory dives elicit a response characteristic of long field dives? In the laboratory the seal does not know how long it will be submerged and so prepares for the worst.

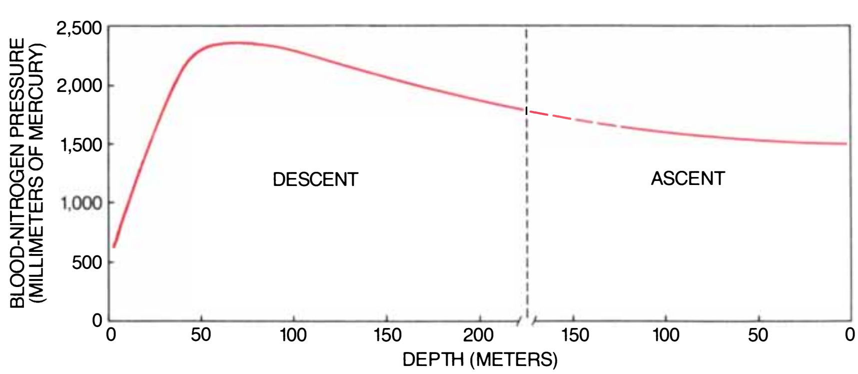
Other studies of oxygen delivery indicate that early in both feeding and exploratory dives at sea the seal actually increases the concentration of red blood cells in the circulation, thereby maximizing the hemoglobin level in the blood and thus the amount of oxygen available to the tissues. In a remarkable discovery Jesper Qvist of Herlev Hospital in Copenhagen, who was a part of our Antarctic team, found that the red-cell concentration in the arteries increases by 50 percent in the first 10 to 15 minutes of a dive. In contrast to my group's laboratory findings, which suggested that the levels are always high, Qvist showed that the cells initially account for only 35 or 40 percent of the circulating blood volume and then rise to 60 percent during a dive. (The levels return to normal within 10 minutes after the animal surfaces.)

Where might the bounty of new cells come from? The spleen is a reasonable guess. This poorly understood organ is known to contract when the sympathetic nervous system is activated, as it is when a mammal dives or is frightened. (Indeed, fear may explain why some seals have elevated red-cell levels when they are confined in a laboratory.) Contraction of the spleen could well inject stored oxygen-rich red cells into the seal's highly expandable venous system; the heart could then deliver them to the arterial circulation as needed. After the seal returned to the surface for air, the circulating red cells would be readily reloaded with oxygen and stored again.

An oxygen-supplying role for the spleen is not unprecedented; the organ is known to infuse red blood cells into the circulation within minutes after a racehorse begins intensive exercise. More direct evidence of the spleen's importance to the seal comes from an-



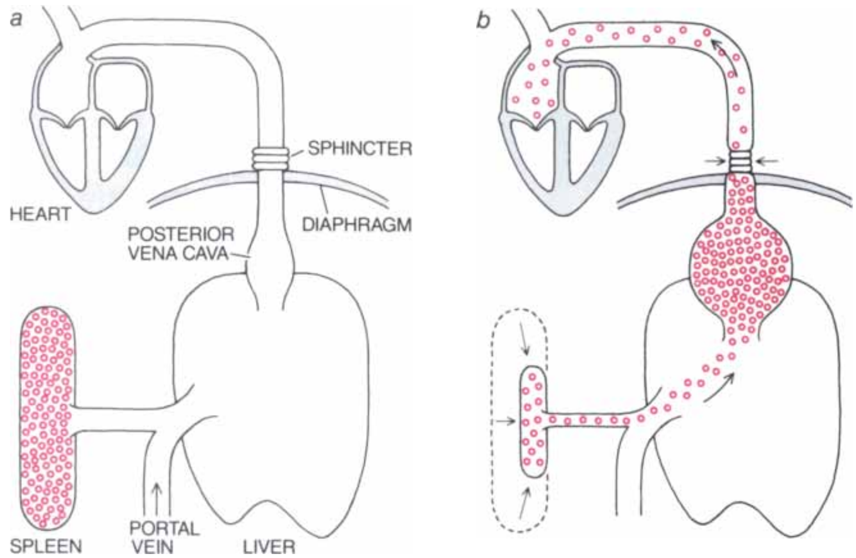
CONCENTRATION OF HEMOGLOBIN (*a*), the oxygen-carrying pigment of red blood cells, rises in the blood during the first 10 to 12 minutes of voluntary ocean dives, preventing the oxygen levels in the seal's blood (*b*) from falling precipitously. In excursions that last for more than 17 minutes (*broken line*), when the muscles switch to anaerobic metabolism, the influx of hemoglobin into the blood actually balances oxygen consumption for about 15 minutes. In trips of less than 17 minutes (*solid line*), when the muscles apparently burn oxygen, the added hemoglobin cannot fully counteract oxygen consumption by tissues; hence the level of oxygen in the blood declines gradually from the start.



NITROGEN PRESSURE in the blood stops increasing when seals reach a depth of 40 meters, indicating that the lungs no longer release gas into the blood; in fact, they collapse. Such collapse limits the amount of nitrogen that can enter the blood during a dive and thereby helps to protect the seal from nitrogen narcosis and the bends. As the dive progresses, uptake of nitrogen by the muscles and the blubber further reduces the risks.

atomical studies of several mammals. On comparing organ weight with body weight I found that the size of the Weddell seal's spleen is particularly large, matched only by that of the southern elephant seal, another long-diving species. In human beings, dogs and even baleen whales the organ is considerably smaller for the animal's bulk. On the basis of organ size and the degree to which hemoglobin levels rise during a dive, my colleagues and I estimate that the Weddell seal warehouses approximately 60 percent of its total red-cell supply in the spleen, whereas a human maintains less than 10 percent there. Indeed, the seal's spleen appears to be something of a contractile scuba tank in its ability to store and release oxygen needed for a dive.

The effects of an oxygenated red-cell infusion become particularly apparent in long dives. In the period when the red-cell concentration is rising, the blood's oxygen content remains constant, indicating that the amount consumed by the brain, heart and other crucial tissues is somehow being replaced. The same plateau is not seen in feeding dives, where the muscles consume oxygen; then oxygen levels fall steadily. In this instance the oxygen burned by the seal's muscles probably outstrips the ability of the splenic blood-storage system to inject red cells into the circulation.



LARGE SPLEEN probably collects oxygen-rich red blood cells when the seal breathes air at the surface (*a*) and injects the cells into the circulation when the animal submerges (*b*). Such activity would explain the finding that hemoglobin levels are elevated at the start of ocean dives. The author suggests that the spleen stores about 24 liters of red blood cells. It contracts when the dive begins, squeezing much of its contents into the portal vein, through the liver and into the expandable posterior vena cava. A sphincter allows the reservoir to release cells to the heart and to the arterial circulation as needed.

In addition to providing oxygen during a dive, the inflow of fresh red blood cells into the circulation probably serves another important purpose: the dilution of gases dissolved in the blood. Such an effect would explain why the carbon dioxide concentration rises surprisingly little in the course of field diving. Dilution would also help to explain why nitrogen does not cause narcosis or the bends in the seal. The field studies suggest other explanations as well. For instance, our international group has found, as others had suggested, that the lung collapses. It does so at about the 40-meter mark, somewhat earlier than predicted.

We determined the point of collapse on the basis of work by Konrad J. Falke of the University of Düsseldorf, who painstakingly measured nitrogen pressures in blood specimens drawn from the arteries of free-diving seals. (Pressure is a good indicator of nitrogen concentration because the tension in the blood rises and falls with the concentration.) Falke found that the blood-nitrogen pressure, which measures 550 millimeters of mercury when the animal breathes at the surface, increases as the seal descends, peaking at from 2,000 to 2,400 millimeters when the animal reaches 40 meters. Beyond 40 meters the pressure falls. We think the drop in pressure that follows the lungs' collapse happens not only because the spleen infuses red cells into the circulation but also because some nitrogen diffuses out of the blood and into muscles and blubber.

As we studied the many remarkable adaptations of the mature seals in McMurdo Sound we became increasingly curious about the responses of the seal fetus. Does it exhibit a diving reflex when the mother descends at sea? Work by Robert Elsner of the University of Alaska at Fairbanks had suggested that it might. He found that the fetal heart rate, like the mother's, slows during laboratory dives.

We do not yet have a complete answer to our question, but we were able to collect some field data when Graham C. Liggins of the National Women's Hospital in Auckland and Hill succeeded in placing a heart-rate monitor on the back of a pregnant seal. The computer record showed that the fetal heart rate does indeed slow during free dives, but the decline is more gradual and less marked than the mother's; the heart rate also accelerates more gradually after the mother surfaces. The fetus "knows" when its parent dives, although exactly what informs it is not clear. We must complete additional studies to determine whether alterations in the fetal heart rate are accompanied by changes in cardiac output and in the distribution of blood flow. If they are, the finding would indicate that the fetus conserves oxygen for vital tissues as the mother dives and replenishes its store at the surface.

As a whole our field studies demonstrate that the Weddell seal's responses to its occasional long dives look much as laboratory dives predict they should. The diving reflex is in full force: the heart rate slows and remains low throughout the dive, and the muscles switch from aerobic to anaerobic metabolism, indicating that their supply of blood is shut off, probably because their arteries are constricted.

In the majority of dives, in contrast, the profile is rather different. When the seal embarks on a feeding excursion, the diving response is modified. The heart rate slows but is more variable, speeding up as the seal swims faster. Moreover, the muscles continue to rely on aerobic metabolism; apparently they continue to receive some blood, indicating that vascular constriction is modulated. Early in its dive the seal apparently "decides" whether its foray will be long or short and whether or not it must resort to draconian measures to conserve oxygen.

The mechanism by which the seal makes the decision is one of many puzzles left to be solved. Given the rapidly advancing technology now available, it may not be long before the question is answered and added insights are gained into the complex adaptations of the Weddell seal, one of the world's most impressive diving machines.

shifting, high-torque 5-speed transaxle turn that raw power into inspired performance.

A NEW LEVEL OF INTERIOR COMFORT.

The Beretta cockpit. A world of aesthetic design. Both front bucket seats have their own suspension systems to help tune out road vibrations. Watch Beretta's vital signs come to light as you turn on the optional electronic instrumentation.

With a touch of your finger, you can even check your gas mileage or the temperature outside.

Make the road your natural habitat; put yourself in the unforgettable shape of

Beretta. A new species.

*Performance figures compiled by a professional driver on a GM test track.

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

GM 6/60 QUALITY COMMITMENT PLAN

See your Chevrolet dealer for terms and conditions of the new limited powertrain warranty.

THE *Heartbeat* OF AMERICA TODAY'S CHEVROLET

The Connection Machine

Most computers have a single processing unit. In this new parallel computer 65,536 processors work on a problem at once. The resulting speed may transform several fields, including artificial intelligence

by W. Daniel Hillis

In the past three decades remarkable changes have taken place in digital computers. The amount of computational power that once required a room full of vacuum tubes can now be found in hand-held devices. Complex computations that would once have taken days to perform can now be done in seconds. Yet in certain fundamental respects the design of the digital computer remained unchanged between the days of the ENIAC (one of the first large-scale digital machines, built at the University of Pennsylvania in the late 1940's) and the current generation of supercomputers. Most modern computers—from supercomputers to microprocessors—are similar to the ENIAC in that the memory and the central processing unit are separate entities. For a computation to be performed, the appropriate data must be retrieved from memory and brought to the central processor; there it is operated on before being returned to memory.

Such a design is called sequential because the processing operations are performed one at a time. The sequential design was adopted mainly for utilitarian reasons. In the early days of digital computing the memory and the central processing unit were made of different materials. Since memory was cheaper than processing, it was desirable to maximize the efficiency of the processing unit at the expense of the memory's efficiency. And that is just what the sequential design does. Today, however, the memory and the central processor are fabricated from the same etched silicon wafers. In a typical computer more than 90 percent of the silicon is devoted to memory. While the central processor is kept wonderfully busy, this vast majority largely sits idle. At about $1 million per square meter, processed packaged silicon is an expensive resource to waste.

Clearly, the general solution to this problem is to find a way to unify processing capacity and memory. But how? One answer is to exploit many small processors, working simultaneously, each accompanied by a small memory of its own. In such a design, which is called parallel processing, memory capacity and processing capacity can both be utilized with high efficiency. This is the approach my colleagues and I have taken in building a parallel computer called the Connection Machine. The Connection Machine incorporates 65,536 simple processors. Each processor is much less powerful than a typical personal computer, but in tandem they can execute several billion instructions per second, a rate that makes the Connection Machine one of the fastest computers ever constructed.

Yet the most interesting thing about the Connection Machine is not its brute speed but its flexibility. Special-purpose devices have been built that exploit parallelism to perform specific tasks quite quickly. Like idiot savants, however, such machines are usually quite awkward outside their specialties. In contrast, the Connection Machine can operate at its peak processing rate in a wide range of applications. As this article will describe, the key to such flexibility is a communications network that enables the multitude of processors to exchange information in the pattern best suited to the problem at hand. The Connection Machine is not just a prototype. About a dozen Connection Machines are already in commercial use, and they have begun to change the way digital computing treats problems in physics, image processing, text retrieval and even artificial intelligence.

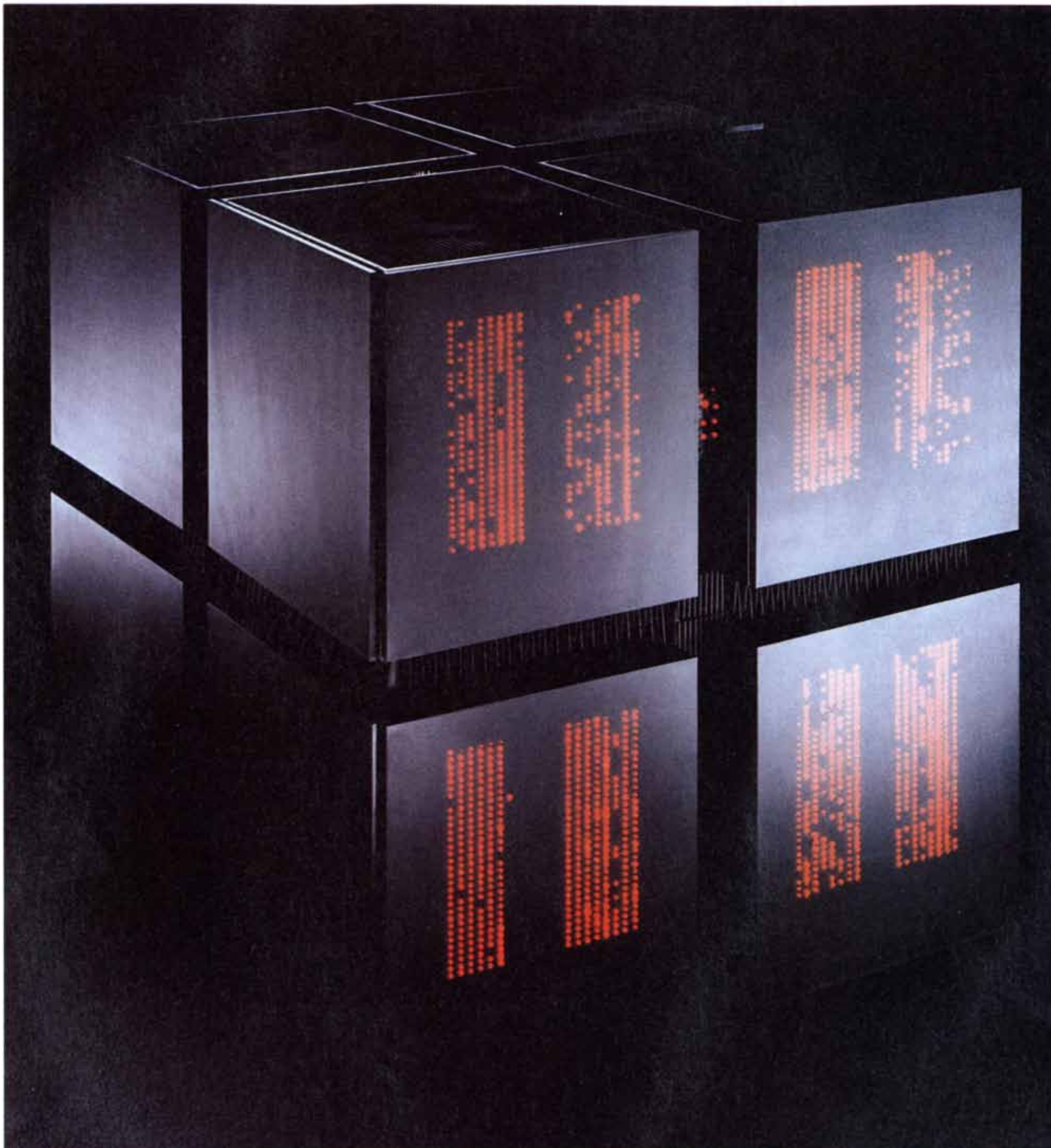
In order to understand the benefits of parallel processing, it is helpful to think about the difference between the way a conventional computer deals with an image and the way the same image is treated in the human brain. From the pair of two-dimensional images falling on the retinas a human being is able—without apparent effort—to reconstruct a three-dimensional model of the world and maintain that model as the two-dimensional images change rapidly. Computers can be programmed to carry out part of the task, but even quite fast computers take hours to do what the human brain can do in fractions of a second [see "Vision by Man and Machine," by Tomaso Poggio; SCIENTIFIC AMERICAN, April, 1984]. The brain maintains its advantage in spite of the fact that its components—neurons—are apparently millions of times slower than the computer's transistors.

Why, then, is the brain so much faster than the computer? The visual circuitry of the brain is not fully understood, but it is clear that in some areas of the brain the principles of parallel processing are at work. In those parts of the brain the entire image is processed at once. The computer, however, examines the image one tiny spot at a time, as if it were looking through a minute keyhole. In the computer the image is represented as an array of numbers, each of which corresponds to the intensity of the light at a particular point. A typical low-resolution array might be a square with 256 points on a side. A conventional computer operates on only one of the square's 65,536 points at a time. Hence even a simple image-processing operation includes 65,536 steps.

The Connection Machine, on the other hand, assigns a single processor to each point of the image. Since every operation can be performed on all the points simultaneously, a calculation involving the entire image is as fast as a calculation involving only a single point. For example, to find all the points in the image that are brighter than a certain minimum a sequential machine must check the 65,536 tiny elements in succession, comparing each one with the threshold value. In the Connection Machine that comparison is made simultaneously by the 65,536 processors—each one operating on a single element of the image.

The threshold comparison is particularly simple because it can be carried out independently by each processor. Most interesting computations, however, require that the processors exchange information as the operation proceeds. Consider the common image-processing operation called convolution. Convolution blurs an image by averaging each point with its nearest neighbors in the two-dimensional grid. (Convolution, which is analogous to operations carried out in the human visual system, is useful for removing insignificant details and bringing out significant objects.)

To complete a convolution each processor must read a value from the processors that store information about the points to the left, right, above and below the point in question.

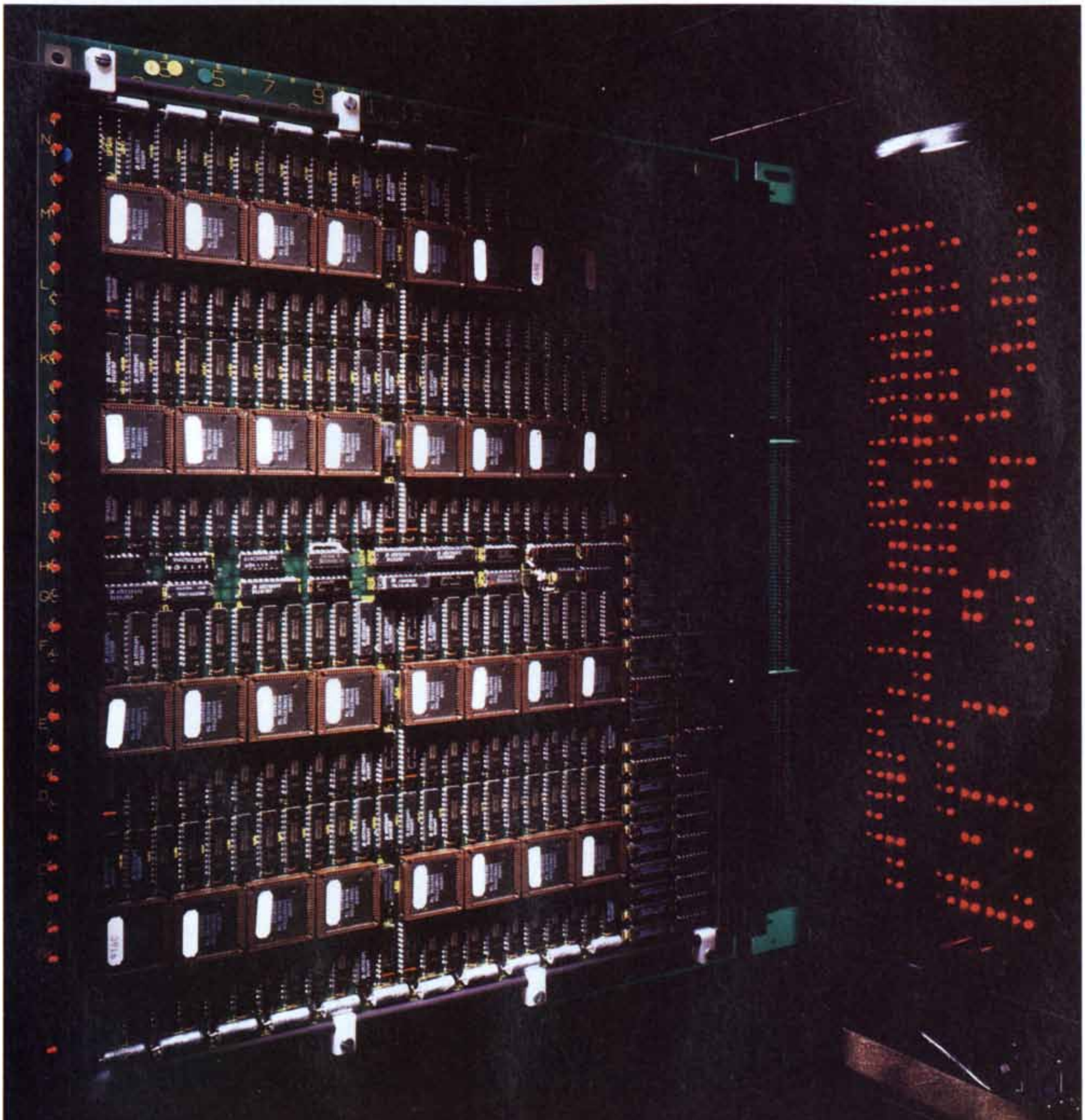


CONNECTION MACHINE is a cube 1.5 meters on a side made up of eight subcubes. Each subcube contains 16 boards arranged vertically. On each board are 32 custom chips. Every chip includes 16 processors, each with a small amount of associated memory. The red lights on the boards indicate the status of the chips; they are for troubleshooting. Operating in parallel, the 65,536 processors can execute several billion instructions per second, making the Connection Machine one of the fastest computers ever built.

In effect the processors must "talk" to each other. One way to accomplish such a pattern of communication is to wire the processors in a two-dimensional grid. Since each processor would be wired to its four nearest neighbors, the grid corresponds directly to the communication paths required for convolution. Indeed, some parallel computers specialized for image processing are wired in a two-dimensional grid. That pattern works well for convolution but not for other computations.

For example, in computing the average intensity of all points in the image a pattern of connections resembling an inverted tree is the most convenient. The average intensity of an image containing 65,536 points can be calculated by first computing the average of every pair of points, then the average of each pair of pairs, and so on. In 16 steps the average can be derived. In its last few steps the computation requires an exchange of information about points that are widely separated in the image; therefore the two-dimensional grid is not a convenient pattern of wiring.

The general principles to be derived from these examples are that each type of computation may require its own pattern of connections and that each processor may need to communicate with any other. Therefore in designing the Connection Machine we chose a communications network in which any processor can communicate with any other. As a result of such flexibility the programmer is free to



BOARD slides out of the Connection Machine much like a book from a shelf. The square objects are the chips, each with its 16 processors. The rectangular objects include memory units and devices for routing communications among the assembled processors.

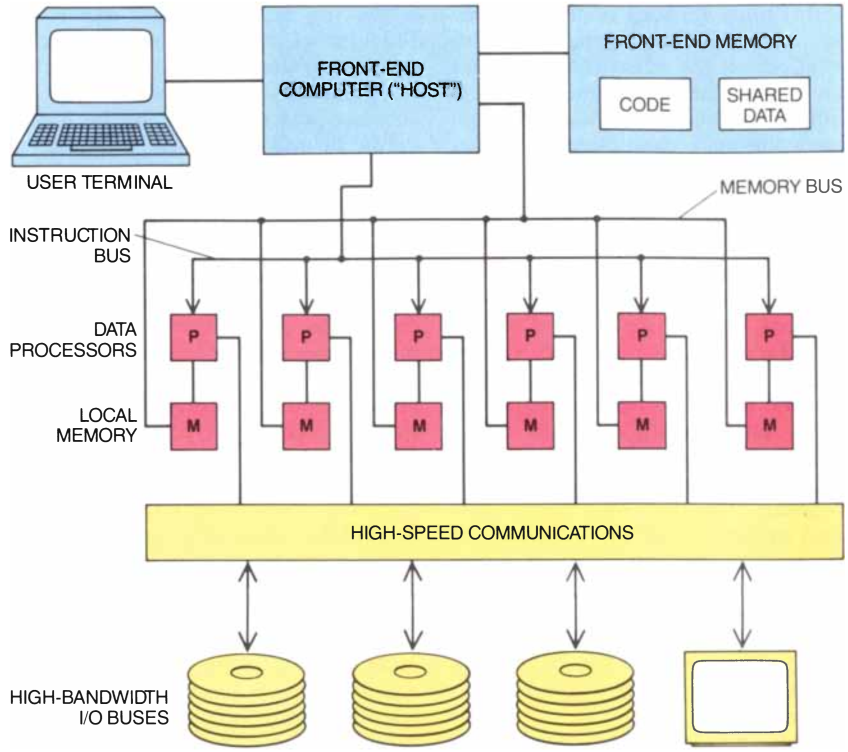
choose the algorithm that is most appropriate for solving the problem at hand without having to worry about the limits imposed by the pattern of wiring.

The basic replicated unit of the Connection Machine is an integrated circuit consisting of 16 small processors and a device for routing communications. Each of the processing units is associated with 4,096 bits of memory. (A typical personal computer has 256,000 bits or more.) The 16 processors are etched on a single chip and 32 of these chips are packaged on a single printed-circuit board. There are 128 such boards in the machine, arranged in a cube 1.5 meters on a side. For purposes of troubleshooting each chip is connected to a light on the edge of its board; the array of lights forms a pattern on the face of the cube as the machine operates.

The 16 processors on each chip are connected by a switch that makes it possible to create a direct connection between any pair of processing units. Implementing such direct connections between every pair of processors among the 65,536 in the system would require more than two billion wires, obviously an impractical figure. Instead, the routing device on each chip is connected to 12 other routers in the system. The routers are wired according to a pattern called a Boolean *n*-cube. The *n*-cube is a generalized version of an ordinary three-dimensional cube that has some excellent properties as a network for communications among processors.

The full mathematical detail of the *n*-cube is somewhat beyond the scope of this article, but its general principles are not difficult to grasp. One can imagine an ordinary three-dimensional cube as one member of a series of "cubes" corresponding to different spatial dimensions. For example, a line segment might be thought of as a "one-cube," or a cube in one dimension. Joining two one-cubes by their ends yields a two-cube, or a square. Joining two two-cubes by their corners yields a three-cube, which is what we ordinarily think of as a cube. Similarly, joining two three-cubes by their corners yields a four-cube [*see illustration on next page*]. The process may be repeated any number of times, and it can readily be shown that a 12-cube has 2^{12} (4,096) corners, or one for each chip in the Connection Machine.

Such a Boolean *n*-cube is a valuable arrangement for several reasons. In the first place, no processor in the 12-cube is more than 12 wires away from any other, which facilitates communication in the network. Second, the design of the *n*-cube accords well with



SYSTEM DIAGRAM shows that the Connection Machine operates in association with a conventional computer, which is called the host. A user of the system interacts with the host by means of a conventional computer language modified for parallel programming. Rather than carrying out repetitive operations one at a time, however, the host delegates them to the Connection Machine, where the operations are done in parallel. The results can be obtained by various input-output devices, including high-resolution visual displays.

the binary logic of the computer. In the digital computer all data are stored as strings of bits, each with a value of either 0 or 1. Now, each cube in the *n*-cube has two subcubes, which may be designated 0 and 1 respectively. As a result each point in the *n*-cube has a unique address specified by a string of 12 binary bits. The first bit specifies which of the 11-cubes within the 12-cube contains the desired point. The second bit specifies which of the 10-cubes is in question, and so on until a unique point has been determined.

These binary addresses can be employed to route messages among the 4,096 chips in the Connection Machine. Each message in the system includes such an address. On receiving the message, the router examines the address one bit at a time, then forwards it to the next router along the way. That router in turn takes up the message, examines the address and forwards it. Thus in no more than 12 steps any message will find its way to the destination.

This communications network has several features that augment its speed and flexibility. One of the valuable properties of the *n*-cube arrangement is that there are many equally efficient routes of communication between any pair of processors. If one route is already occupied by a transmission in progress, the router is free to select an alternate route merely by processing the bits of the address in a different order [*see illustration on page 113*].

Another type of flexibility is also inherent in the communications system. In some instances the communications network behaves more or less like a telephone exchange: it establishes a circuit between two processors so that they can communicate continuously and exclusively. In complicated cases, however, the messages may be so long and the system so crowded that the routers must behave more like post offices, storing packets of information that are later forwarded. Such decisions are made by the routers based on what wires are available when a transmission must be made.

These properties make it possible for the Connection Machine to establish many different patterns of communication, depending on the problem at hand. An important feature of the system is that such details are invisible to the user, who needs to know no more about Boolean *n*-cubes than the average user of the telephone needs to know about digital switching. (Indeed, future versions of the machine may have other wiring patterns with no effect on the algorithms that are employed.) The programmer interacts with the Connection Machine through a conventional computer, known as the host, which employs a standard operating system and programming language. The processors of the Connection Machine are connected with the host much as a conventional memory unit would be.

Indeed, in one sense the Connection Machine is the memory of the host. That relationship makes possible a simple integration of parallel computing and existing software. Programs for the Connection Machine are surprisingly similar to conventional programs. The chief difference is that many operations normally carried out by repetitive loops are replaced by single operations corresponding to the simultaneous operation of many processors in the Connection Machine; the routing hardware automatically establishes the necessary communication paths.

It should be noted that nowhere in this system is exotic hardware to be found. In designing the Connection Machine we chose well-tested technologies in order to achieve simplicity and reliability. The individual processors are relatively slow by the standards of today's fastest computers. The custom chip is built by methods similar to those for making personal computers and pocket calculators. Yet the assembled power of the 65,536 processors makes the machine very fast. For many applications the machine can perform more than two billion operations per second; for the most favorable applications the figure is more than 10 billion, or about 1,000 times as fast as a typical mainframe computer.

Putting the machine's speed in a slightly different context, one might consider floating-point operations, which provide a standard for computing power in number-intensive scientific applications. A floating-point operation is the multiplication or addition of two numbers expressed in scientific notation (such as 1.5×10^2). A typical supercomputer can carry out a few hundred million floating-point operations per second; the Connection Machine can average about 2,500 million on a typical problem.

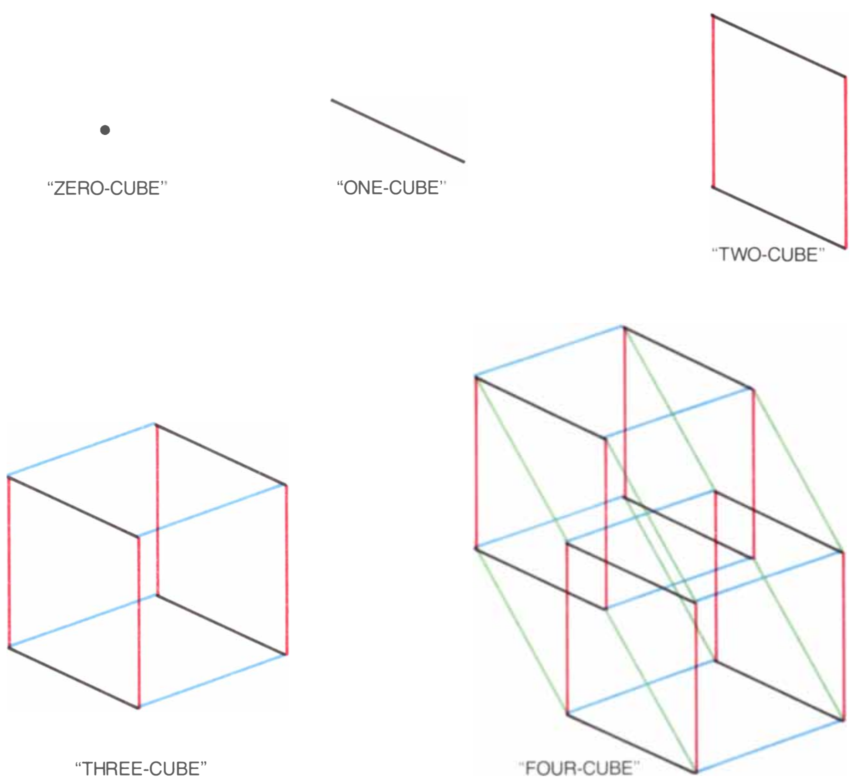
To what uses has this considerable number-crunching capacity been put? As suggested above, some of the initial applications have involved manipulating and processing images. Others have exploited the parallelism inherent in certain physical processes. The engineering problem of calculating the flow of air over an airplane wing or a helicopter rotor provides an example of how the Connection Machine can mirror the parallelism of nature.

In nature the overall flow pattern emerges from the myriad interactions among air molecules, which bump into one another and into the surface of the wing as they rush along. The engineer (who is interested in the overall flow rather than the specific molecular interactions) uses a simplified, large-scale model consisting of a set of partial differential equations. Yet the equations themselves are set up in parallel: they treat changes in pressure in small volumes of air and sum their interactions to yield the overall flow. Because the equations are parallel, their solution is fast and efficient on the Connection Machine.

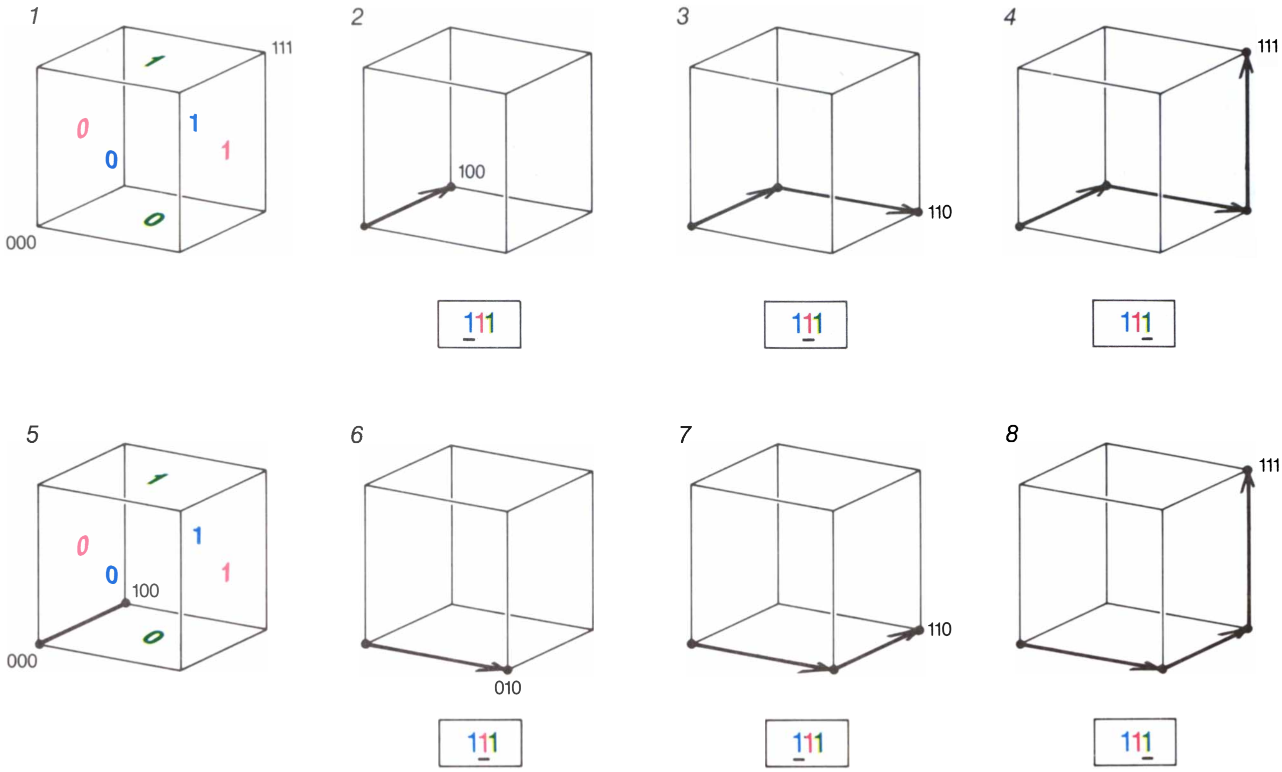
With a parallel computer, however, one can also move beyond the equations and come closer to the underlying physical reality. The large-scale behavior of a fluid is for the most part independent of the detailed physical properties of its individual particles. Moreover, the qualitative behavior of the fluid is not changed when the number of particles is greatly reduced. Therefore it is possible to accurately re-create large-scale flows by examining collisions among a few tens of millions of simple, generalized particles.

Stephen Wolfram of the Center for Complex Systems Research at the University of Illinois at Urbana-Champaign and my colleague James Salem took advantage of this technique to model fluid flows over complex surfaces. Their simulation entailed only a few tens of millions of particles, and the particles were allowed to move in only six directions at integral velocities. Nonetheless, the system is capable of accurately mimicking the flow of a fluid.

The simplest and most logical way to perform the fluid-flow computation would be to assign each particle its own processor. Yet a typical simulation includes about eight million "particles," and the Connection Machine, vast as it is, includes only some 65,000 processors. The solution to this programming difficulty and analogous ones is to program each processor to act as if it were a string of different



BOOLEAN *N*-CUBE provides the topology for the network that links the Connection Machine's processors. A Boolean *n*-cube is a generalized version of an ordinary cube. Such cubes can be constructed in many dimensions, each building on the next-lower dimension. A point can be considered a cube in zero dimensions, or a "zero-cube." Linking two points yields a "one-cube," or a line. Linking a pair of "two-cubes" (squares) yields the familiar three-dimensional cube. Two three-cubes can be joined at their vertexes (corners) to form a "four-cube." Repeating the process would yield a "12-cube" with 4,096 vertexes. The 4,096 chips of the Connection Machine are wired in the form of a 12-cube.



ALTERNATE ROUTES for communication between chips are provided by an *n*-cube. The illustration shows alternate routes in a three-cube, but the same principle applies to the 12-cube of the Connection Machine. Each vertex of the *n*-cube (where the chips lie) can be assigned a unique address as follows. A three-cube includes three pairs of planes. Each plane can be designated 0 or 1, and a vertex is then assigned a three-digit address according to which member of each pair of planes it is found in (*1*). Messages are forwarded by routing devices at each vertex, which read the address and process it one digit at a time. Here a message is sent from 000 to 111. The router reads the first digit and forwards the message to point 100 (*2*). There the second digit is read (*3*). At 110 the third digit is read and the message is forwarded to its destination (*4*). When it comes time to send the message, however, the wire between 000 and 100 may be busy (*5*). In that case the router simply reads the second digit of the address first, choosing an alternate route (*6*). Then the first and third digits are read (*7, 8*) and the message is delivered to the correct address.

processing units, each unit handling one particle at a time. The details of the arrangement are again invisible to the programmer, who simply specifies how many "virtual processors" are required. The hardware and software take care of the rest. Of course, if each processor must simulate 250 units in turn, the computation takes 250 times as long as it would with one actual processor per particle.

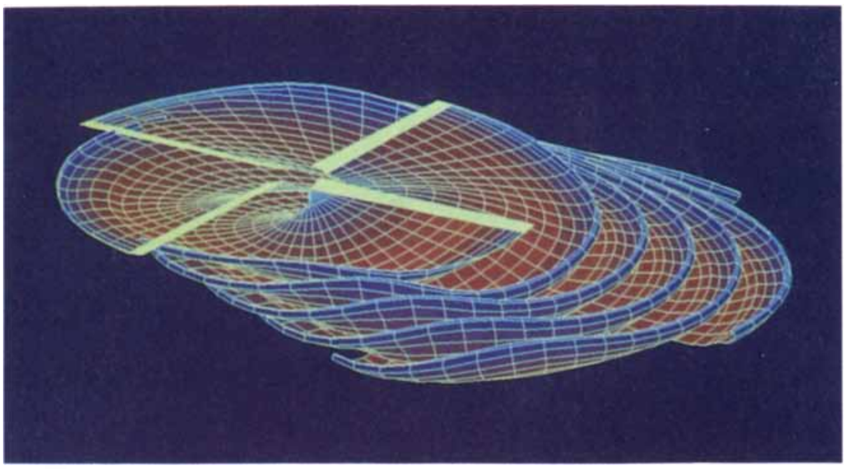
Many interesting applications of the Connection Machine do not involve numbers. My colleagues Brewster Kahle, Craig Stanfill and David Waltz exploited the computer's parallelism to retrieve documents from large collections of texts. The underlying principle of their system is that each processor can be programmed to compare one document in a large data base with a "search sample," a paragraph chosen for its relevance. Once the comparison has been made, the processors exchange information and rank the documents according to how well they match the search sample.

Comparing two pieces of prose to see how well they match is not a simple task. Merely counting the number of words that appear in both samples is useless, because the count is contaminated by words such as "the" and "as," which carry little content. Therefore the document-retrieval system exploits a dictionary and some rules of grammar to extract from each sample the phrases that bear its content. Each processor is loaded with a different article compressed in this way, and the search sample is broadcast to all the processors in the network.

The process of comparison is relatively simple, and since 65,536 documents are checked at once, the entire data base can be examined almost instantaneously. Ranking the articles according to how well they match the search sample is a more difficult operation, because it requires a complex pattern of communication among the processors. Yet in parallel it can be performed in about 50 milliseconds. A few of the highest-ranked documents are then offered to the user of the system, who can choose a new search sample from among them. (Conversely, articles that are clearly irrelevant can be chosen as negative search samples.) Because all the comparisons are done at once, the entire collection of texts can be winnowed repeatedly in a short period, which ensures that all relevant articles will be found.

The document-retrieval program is able to function without anything that approaches an understanding of the contents of the articles. Actually understanding those contents would require considerable background knowledge about the world, which has not been incorporated into the retrieval system. One exciting area of research involving the Connection Machine is the writing of programs that include such background knowledge and are able to mimic certain aspects of human reason.

Like the processing of two-dimensional images to form a three-dimensional world model, "commonsense" reasoning is carried out without apparent effort in the human brain. For example, any child can deduce (with the appropriate affective response) that his mother's favorite vase will fall if it is dropped. The child is able to infer that the vase is more like a plate or a rock, which fall, than it is like a bird



HELICOPTER ROTOR produces a complex airflow that can readily be simulated in parallel on the Connection Machine. Each processor models the circulation within a certain layer of air (*small subdivisions of the image*). The circulation of the air in each section influences the air in each of the other sections. These interactions are computed in parallel. Details such as the wavy distortion at the bottom are important for predicting forces exerted on the helicopter blades. The simulation was developed by T. Alan Egolf of the United Technologies Research Center and the author's colleague J. P. Massar.

or a balloon, which do not. The inference can be made correctly in spite of apparently contradictory information, such as the fact that the vase may be spherical, as a balloon is, or that the child's mother may also own a bird.

Clearly, for human beings the ease and accuracy of such inferences increases with the accumulation of knowledge about the world. The opposite is true for conventional computers. As the number of different concepts increases, the number of possible relations among them increases even more quickly. Because a sequential computer can examine these relations only one at a time, its pace slows dramatically as the quantity of background data grows. Indeed, the slowness of conventional computers in commonsense reasoning was one of the stimuli responsible for the design of the Connection Machine.

In the late 1970's, as a graduate student at the Massachusetts Institute of Technology, I became interested in how commonsense reasoning might be simulated by computers. It seemed to me that one way out of the morass sequential computers found themselves in when asked to make simple, everyday deductions was to build a machine that could examine possible connections among concepts more than one at a time. (In that conclusion I was inspired by the work of Marvin L. Minsky of M.I.T. and Scott E. Fahlman of Carnegie-Mellon University.) That was in 1978. By 1985 the idea had moved to the stage of an actual prototype with the aid of a grant from the U.S. Defense Advanced Research Projects Agency, which offered to buy the first machine. By then I had left M.I.T. and helped found the company—Thinking Machines Corporation—that builds and markets the Connection Machine.

Now that the Connection Machine is a physical reality, investigators of artificial intelligence are making use of it to solve commonsense reasoning problems. The germ of this approach is to assign one fundamental concept to each processor. The connections among processors can then be exploited to represent multiple relations among simple concepts. In the simple example given above, one processor may represent the concept "Vase," another "Mother" and a third "Likes." The connections among these three processors would embody the knowledge that "Mother likes her vase." Other connections might represent the vase's shape, composition and history. When it comes time to decide what the outcome would be if the vase were dropped, the relevant connections can be searched in parallel.

Since there are now about a dozen Connection Machines in operation, there will undoubtedly soon be many new programs for the machine. It is likely that many of them will be in the four general areas touched on above: image processing, simulation of physical processes, searching of data bases and artificial intelligence. One of the greatest challenges in learning to use the Connection Machine lies in beginning to think in parallel terms. Programmers have considerable accumulated experience in programming for sequential machines, and such programming has by now become almost second nature. Learning to write programs for parallel machines requires thinking in ways that are quite different from those demanded by sequential computers.

That challenge will be greater for the Connection Machine than it will be for some other types of parallel machines. It should not be assumed that the Connection Machine is the only representative of its genre. Indeed, many different parallel designs are now in various stages of realization. To generalize greatly, these designs fall into two broad classes: "coarse-grained" and "fine-grained." Coarse-grained machines link relatively few processors, each with a relatively large amount of computational power; fine-grained machines link a great many weak processors.

These two classes of parallel computers form a spectrum. At one end is the conventional sequential computer, which has the minimum number of processors: one. At the other end of the spectrum are designs such as that of the Connection Machine, which include a very large number of small processors. Although some highly qualified investigators and companies are pursuing the coarse-grained approach, I think it is the fine-grained design that will ultimately prove the most fruitful. Yet it is also the one that is the most foreign to our preconceptions about computer programming.

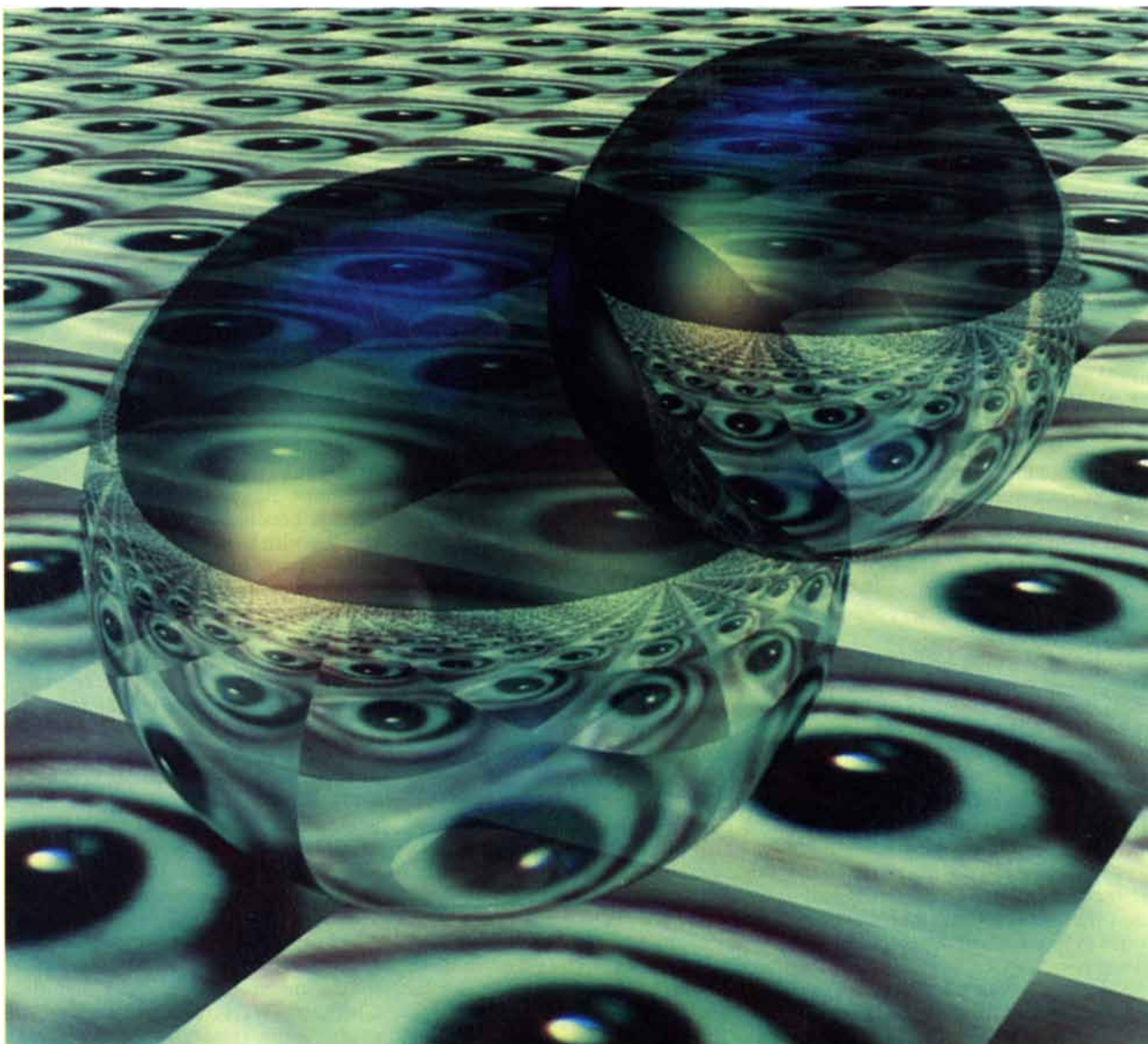
In writing a program for a coarse-grained machine, one can adhere to concepts much like those used for programming sequential computers; the problems arise in attempting to coordinate the programs. In writing a program for the Connection Machine, however, one is faced with an entirely different realm of problems and possibilities. Exploiting the full potential of the machine will require a new way of thinking about computation, which we as programmers have just begun to learn. That learning process will undoubtedly be both rewarding and challenging.

Some of its rewards may come from the fact that the Connection Machine can be expanded to encompass considerably more computational power without any fundamental changes in design. Most of the applications envisioned for the machine could profitably exploit a computer much larger than current versions of the Connection Machine. For this reason the computer has been designed to allow a significant increase in the number of processors. The Connection Machine can be expanded simply by adding processors, memory and communication devices to an existing machine.

As an extreme example of scaling up, imagine a parallel computer with one billion processors. Such a machine might well incorporate some features of the Connection Machine, although there would undoubtedly be many new problems to solve. If built with current technology, a billion-processor machine would be as large as a building and cost 20 times as much as today's largest computers. It could, however, execute some 100 million million (10^{14}) instructions per second, which is several orders of magnitude greater than the computational power of all existing supercomputers combined.

There are technical problems inherent in building such a computational engine, but they are soluble. The real problems are those of the imagination: conceiving how such power would be used. Some engineering problems, including extrapolations of examples mentioned above, might benefit from such capacity, but they are in a sense trivial. The applications worthy of a billion-processor machine are those that entail a radical change in the way we think about computation.

A parallel computer with a billion processors might provide the basis for a computational utility analogous to existing gas and electric utilities. Just as a coal-fired plant generates electricity that is transmitted to individual appliances, a huge parallel computer could provide computational power to a city's worth of robots and workstations. The design of the parallel machine would enable many users to draw on portions of the total computing capacity for small problems, whereas the total capacity could be applied to large ones. Such a vision is somewhat utopian (at least for the moment), but it is by no means impracticable, which suggests the depth of the changes that parallel computing may ultimately bring.



COMPUTER GRAPHICS is one of the fields in which parallel computers may be most fruitful. The illustration was made by the technique called ray tracing, in which each Connection Machine processor is assigned to a different pixel (picture element). The processors trace rays of light bouncing among imaginary objects, here glass balls and images of eyes. The paths of the rays determine the final color of each pixel. Karl Sims of the Massachusetts Institute of Technology Media Laboratory generated the image.

The Birth of the U.S. Biological-Warfare Program

Recently declassified Government files reveal the events that led to research on biological weapons. Now a divisive public issue, the program started out as an obscure operation in World War II

by Barton J. Bernstein

"Why is it so confidential to destroy insect pests?" a perplexed Franklin D. Roosevelt asked his special assistant Wayne Coy on July 14, 1943, when the U.S. had been embroiled in World War II for a year and a half. The chief executive was looking over a Department of Agriculture request for $405,000 to support research on insect infestations and plant diseases. He knew only that his Bureau of the Budget was not informed about this project and that the War Department had instructed Agriculture to keep it secret.

Roosevelt asked Coy to investigate the mysterious enterprise. Two days later Coy told the president that details of the project were guarded for good reasons and suggested he ask a man named George W. Merck to tell him more about the work. Merck was in charge of a civilian adjunct to the War Department. Its mission was research on biological warfare.

Compared with the $2-billion Manhattan Project that gave rise to the atomic bomb, U.S. research on biological warfare during World War II was a small-scale venture: the project had a staff of about 4,000 workers, including scientists, and its total cost was about $60 million, including construction. Perhaps, then, it is not surprising that President Roosevelt did not remember that the Agriculture Department had joined the Army's Chemical Warfare Service in its efforts to develop biological weapons.

Neither Roosevelt nor Harry S Truman, his successor, ever confronted the decision to order a biological attack. The issue of biological warfare did not end, however, with World War II. The U.S. continued to develop its biological arsenal for many years after the war and American interest in funding such research has recently revived: appropriations for biological-weapons projects, which hit rock bottom during the Nixon, Ford and Carter administrations, climbed back to $60 million last year.

Last September a multinational congress met to review a 1972 convention that bans the development, production and possesion of biological weapons, including toxins, for offensive purposes. Both the U.S. and the U.S.S.R. are signatories of the convention, but the U.S. has charged the Soviets with several violations. Following up on informal contacts, scientific contingents from both countries gathered at Geneva in April to try to hammer out mutually acceptable measures for treaty verification. Given the controversial American allegations, Soviet reluctance at disclosure and disputes on each side about the other's activity, the 1972 agreement could be in trouble.

Under these circumstances it may be instructive to study the origins of the program and examine the decisions that stayed the use of biological weapons by the U.S. in World War II. About 100 key documents have helped me to piece together an account of the wartime deliberations. The documents were culled from thousands of American papers declassified on special request and from British files that were once secret.

In World War I chemical agents such as chlorine and mustard gas killed or injured more than a million soldiers and civilians. Outrage at these deaths prompted 40 nations in 1925 to sign the Geneva Protocol, which prohibited the first use of chemical and biological weapons but placed no constraints on research, production and stockpiling. In subsequent years most major industrial powers maintained active development programs. Although the U.S. signed the Geneva Protocol, it did not ratify the treaty until 1975.

The U.S. Army started conducting biological-warfare research in 1941 through its Chemical Warfare Service, but American efforts did not become substantial until 1942. In February of that year a special committee appointed by the National Academy of Sciences submitted a report to Secretary of War Henry L. Stimson containing recommendations for the future of the biological-warfare program. Stimson had requested the report a few months before the bombing of Pearl Harbor.

The committee, composed of eminent biologists such as Edwin B. Fred of the University of Wisconsin and Stanhope Bayne-Jones of Yale University, concluded that an enemy attacking with biological weapons could gravely harm human beings, crops and livestock. Although the report stressed defense and called for work on vaccines and protection of the water supply, the committee also recommended, rather vaguely, that the U.S. conduct research on the offensive potential of bacterial weapons.

Spurred by the scientists' warnings, Stimson sought presidential approval for a formal biological-warfare program that would include a small group of advisers to coordinate and direct all Government research. "We must be prepared," Stimson wrote to Roosevelt in an April 1942 memorandum. "And the matter must be handled with great secrecy as well as great vigor."

Stimson never mentioned that the Chemical Warfare Service had already begun research into biological weaponry, and the president probably did not know of the program. Still, the chemical service later received millions of dollars in appropriations through the Army's budget and became more instrumental in the biological-warfare program than the small advisory group that directed it. Why did Stimson press for the group?

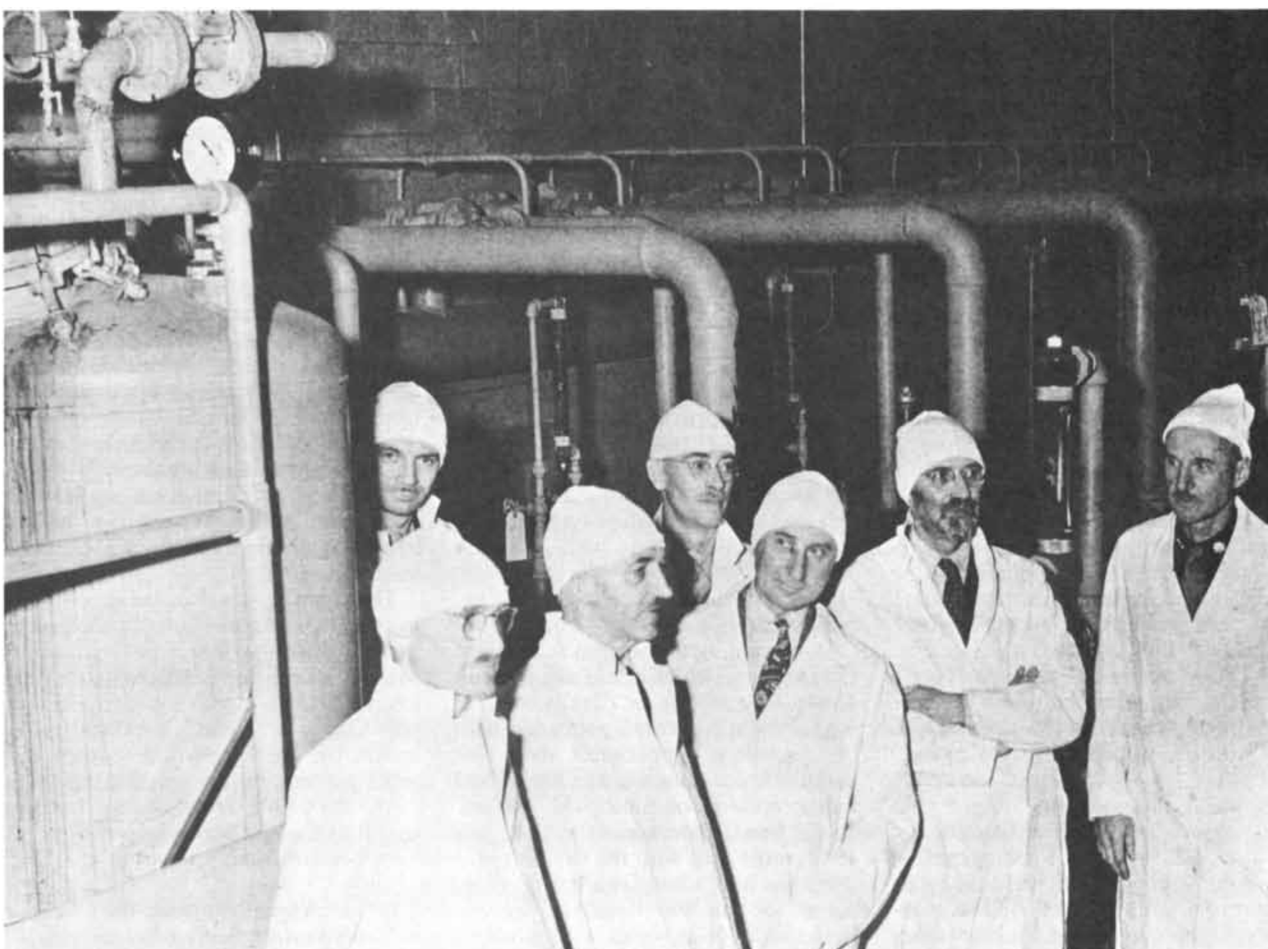
Perhaps it was because, as he told Roosevelt, "biological warfare is dirty business." Stimson hoped to legitimize the research at the Chemical Warfare Service by naming civilians as monitors. Whereas some members of the National Academy of Sciences committee thought the program should be administered by the War Department, top Army officials preferred the establishment of a civilian agency with ties to the armed services. Stimson explained their reasoning to Roosevelt: "Entrusting the matter to a civilian agency would help in preventing the public from being unduly exercised over any ideas that the War Department might be contemplating the use of this weapon offensively."

Stimson suggested hiding the "germ warfare" advisory group in a New Deal welfare agency, called the Federal Security Agency, that oversaw the Public Health Service and Social Security. He wanted an academic luminary to direct the program, someone familiar with the university research system and skilled in administration. After a cabinet meeting on May 15 Roosevelt admitted he had not yet read the secretary's plan but told him to go ahead with it anyway. A week later Stimson discussed his ideas with Secretary of Agriculture Claude R. Wickard, whose agency would later take part in the research coordinated by the advisory group, and with Paul V. McNutt, who led the Federal Security Agency.

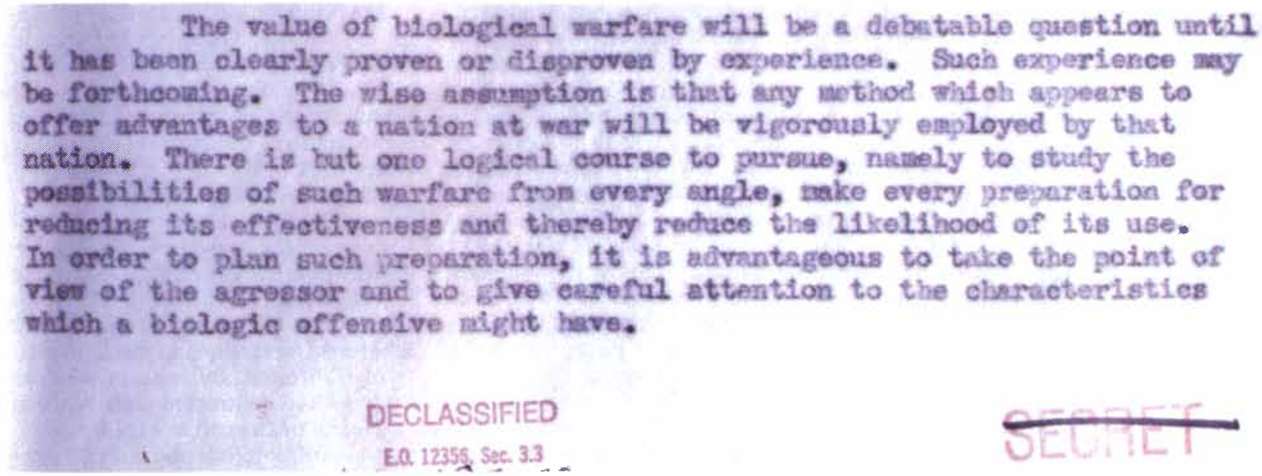
By midsummer three candidates had rejected an offer to head the new group: economist Walter W. Stewart, who chaired the Rockefeller Foundation, geographer Isaiah Bowman, president of Johns Hopkins University, and economist Edmund Ezra Day, president of Cornell University. Finally, in August, chemist George W. Merck, president of the pharmaceutical firm Merck & Co., Inc., accepted the position.

The innocuously named War Research Service (WRS) started out in mid-1942 with an initial allocation of $200,000. Wide contacts among major biologists and physicians enabled the eight-member directorate to initiate secret work in about 28 American universities, including Harvard University, Columbia University, Cornell University, the University of Chicago, Northwestern University, Ohio State University, the University of Notre Dame, the University of Wisconsin, Stanford University and the University of California. By January of 1943 the WRS had contracted with William A. Hagan of Cornell to explore offensive uses of botulism and with J. Howard Mueller of the Harvard Medical School to study anthrax.

Anthrax and botulism remained the foci of biological-warfare research during the war. Both deadly diseases are of bacterial origin, and the bacteria are hardy and prolific. Both have very short incubation periods, lasting for only a few days or even hours. The



TEAM OF SCIENTISTS inspects the facilities at Camp Detrick. In 1943 the camp was the center of the Chemical Warfare Service's biological-weapons program. Pictured are (*front, from left*) N. Paul Hudson of Ohio State University, Guilford B. Reed of Queen's University, Charles A. Mitchell of the Dominion Department of Agriculture, Everitt G. D. Murray of McGill University and Col. Oram C. Woolpert; behind them are James Craigie (*right*) of the University of Toronto and Col. Arvo T. Thompson.


The value of biological warfare will be a debatable question until it has been clearly proven or disproven by experience. Such experience may be forthcoming. The wise assumption is that any method which appears to offer advantages to a nation at war will be vigorously employed by that nation. There is but one logical course to pursue, namely to study the possibilities of such warfare from every angle, make every preparation for reducing its effectiveness and thereby reduce the likelihood of its use. In order to plan such preparation, it is advantageous to take the point of view of the agressor and to give careful attention to the characteristics which a biologic offensive might have.

DECLASSIFIED

E.O. 12356, Sec. 3.3

SECRET

EXCERPT FROM RECOMMENDATIONS made in a 1942 National Academy of Sciences report stresses the feasibility of conducting research on biological weapons. The report probably persuaded President Roosevelt to create the War Research Service.

tough but virulent anthrax spores can be inhaled or absorbed through breaks in the skin; botulism results from ingestion of the bacterial poison botulin.

Reaching beyond college campuses, the WRS empowered the Chemical Warfare Service to expand greatly its own work on biological warfare. In 1942 and 1943 the chemical service received millions of dollars to build research facilities. The most notable one was Camp Detrick in Frederick, Md. (now Fort Detrick), which cost nearly $13 million. The service also hired many scientists to work there and elsewhere in the newly enlarged system.

The scientists, drawn largely from university faculties, put aside their repugnance at developing agents of death because the work seemed necessary in the exceptional situation of World War II. Theodor Rosebury, a Columbia microbiologist, argued in early 1942 that "the likelihood that bacterial warfare will be used against us will surely be increased if an enemy suspects that we are unprepared to meet it and return blow for blow." Soon afterward Rosebury entered the Chemical Warfare Service's laboratory and became a leader at Camp Detrick. "We were fighting a fire [the Axis]," he later wrote, "and it seemed necessary to risk getting dirty as well as burnt."

Stimson and McNutt might well have applauded these sentiments, but they would have been astonished at Rosebury's view of who held the reins. Rosebury believed the ethical concerns of the scientists in his laboratory governed the use of the weapons they were creating. He wrote years later: "Civilians, in or out of uniform, made all the important decisions; the professional military kept out of the way. We resolved the ethical question just as other equally good men resolved the same question at Oak Ridge and Hanford and Chicago and Los Alamos."

History tells a different story. Even though the president himself may not have set the course of the WRS, it seems clear that the key decisions were made in Washington, not in the laboratory.

In spite of Paul McNutt's primary concern with welfare and social services, he kept an eye on the secret biological-warfare program hidden in his agency. In February of 1943 McNutt informed President Roosevelt that the last of the WRS's $200,000 was being spent. The president, he said, would have to decide whether to "go more deeply into two or three...projects now under way." By April, with Stimson's approval, McNutt requested another $25,000 for the WRS fiscal 1943 budget and a total of $350,000 for fiscal 1944. Two days later Roosevelt endorsed McNutt's request with one laconic notation: "O.K. F.D.R." The WRS 1944 budget grew again several months later, when Roosevelt expanded it to $460,000.

In keeping with the tight security of the program, McNutt did not commit particular projects or details to writing, even in his correspondence with the president. Roosevelt's own files contain fewer than a dozen letters and memorandums on biological warfare. Of the handful pertaining to 1942 and 1943, most deal with the small appropriations and administrative arrangements for the War Research Service. Perhaps in discussions with McNutt and Stimson or in meetings with Gen. George C. Marshall, the trusted Army chief of staff, Roosevelt was kept informed of the additional millions of dollars in appropriations going to the biological-warfare work of the Chemical Warfare Service. Not one of the available records, however, shows that Roosevelt was receiving such reports.

Meanwhile the chemical service was enlarging its facilities for development, testing and production. In addition to the 500-acre Camp Detrick site, a 2,000-acre installation for field trials was established on Horn Island in Pascagoula, Miss. A 250-square-mile site near the Dugway Proving Ground in Utah was designated for bombing tests and 6,100 acres were secured for a manufacturing plant to be built near Terre Haute, Ind.

The technology was also advancing. With British technical assistance, the chemical service gained considerable ground in making biological bombs and in late 1943 began work on 500-pound anthrax bombs. These bombs held 106 four-pound "bomblets" that would disperse and break on impact. The bombs were untested, but it was known that pulmonary anthrax, which causes lesions on the lungs, was almost invariably fatal.

The chemical service also succeeded in producing botulin, one of the most potent of all gastrointestinal poisons. Merely tasting food infected with the toxin is usually sufficient to cause severe illness or death. In natural outbreaks the death rate ranges from 16 to 82 percent, but by varying the toxin and the delivery mechanism, the scientists at Camp Detrick hoped to produce a reliably lethal weapon.

Bolstered by its progress, the Chemical Warfare Service began lobbying early in 1944 for an additional $2.5 million to finance the manufacture of anthrax and botulin bombs. The service could produce either 275,000 botulin bombs or one million anthrax bombs every month with that allocation, but it would need time to build factories. Hence the weapons would not be available in quantity until 1945, by which time, military strategists predicted, only the war with Japan would remain.

The service got its funds. Although the vision of a biological arsenal with which to confront the Japanese was tempting, a more urgent threat may have underscored the significance of the service's research. Early in 1944 Allied intelligence experts were beginning to fear that Germany's powerful new V-1 "buzz bombs" might soon be directed against Britain or the troops in Normandy, and that the missiles' warheads might be loaded with germ-warfare agents. The German high command, the experts warned, was facing a strategic crisis; it was assembling all its resources and might resort to biological warfare to gain a permanent advantage.

The analyses were based on so-called worst-case assumptions. They were not comforting; by June, 1944, the U.S. had probably prepared only a few anthrax bombs for testing, if any. Certainly no bombs were available for use against an enemy.

To deter Germany from launching a biological strike, military leaders arranged to inoculate about 100,000 soldiers against botulin, hoping to convince the Germans that Allied troops were preparing for biological retaliation. If Germany had actually staged a biological attack, Anglo-American forces would probably have retaliated with gas.

Germany never called the bluff. Hitler used only conventional explosives in the V-1. As a matter of fact, for reasons that are still not known he had barred all research on offensive biological warfare. The American program—developed substantially to deal with a German threat that never existed—remained untried.

Work at Camp Detrick moved at a brisk pace. In May, 1944, Stimson and McNutt presented Roosevelt with a brief research summary that allotted only five lines to scientific developments. Much more could have been said. An anthrax plant received authorization through the Chemical Warfare Service to manufacture a million bombs and the service was making headway with short-range dispersal techniques for botulin in paste form. In November, Merck sent a report to Stimson and Marshall—but not to Roosevelt—that cryptically mentioned research on four additional "agents against men." Judging from other sources, these were probably brucellosis (undulant fever), psittacosis (parrot fever), tularemia (rabbit fever) and the respiratory disease glanders.

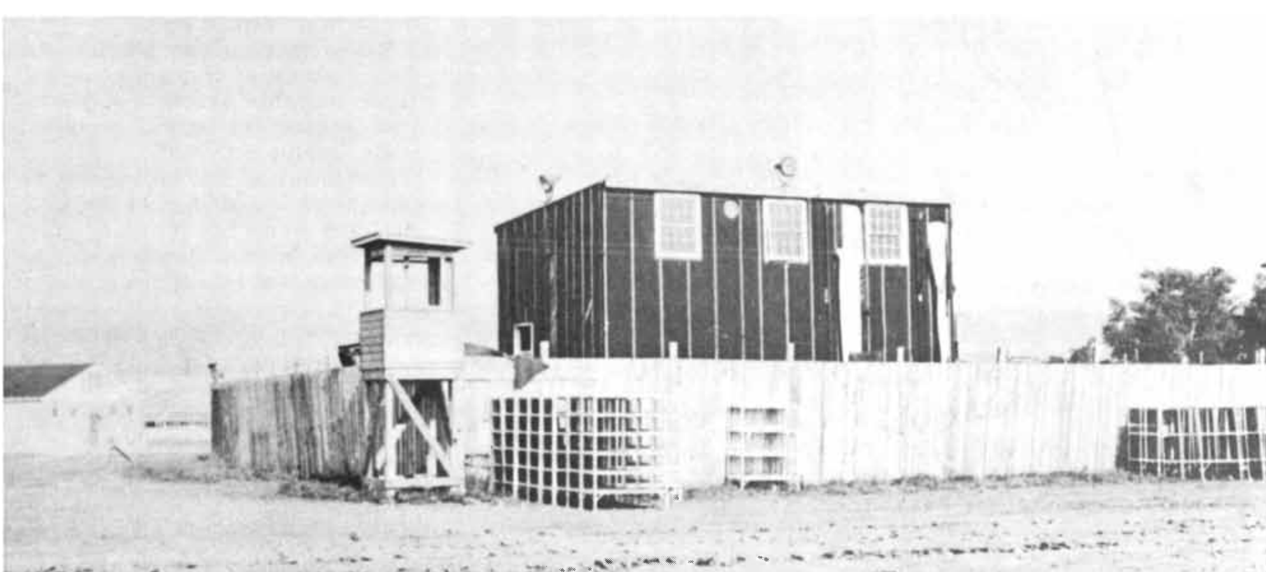
Merck said the Chemical Warfare Service was also developing "at least five agents for use against plants." (These agents are actually chemicals, but at the time they were defined as part of the biological program because they could kill crops.) A sixth compound, ammonium thiocyanate, was recommended for the decimation of "Japanese gardens."

These developments constituted 12 lines in Merck's short November report on biological warfare. The document is tucked away in Stimson's declassified Secretary of War records in Washington. There is no evidence that the secretary or the president devoted any attention to the scientific grist of the program.

Roosevelt neglected not only the science but also the politics of biological warfare. In spite of the considerable progress at Camp Detrick and fears of a German biological offensive, the president seems to have given the matter of biological warfare little thought. In 1942 and again in 1943 Roosevelt had promised publicly not to initiate gas warfare, but he threatened retaliation in kind if the Axis used gas. Apparently he never considered issuing a similar statement on germ warfare. Nor did any adviser propose such a warning to deter action by Germany or Japan.

In May, 1944, Roosevelt's ties to the biological-warfare program became even more tenuous when Stimson and McNutt urged him to abolish the War Research Service and make Merck a consultant to Stimson. The president readily acceded to this reorganization, which may have further distanced him from the secretive enterprise.

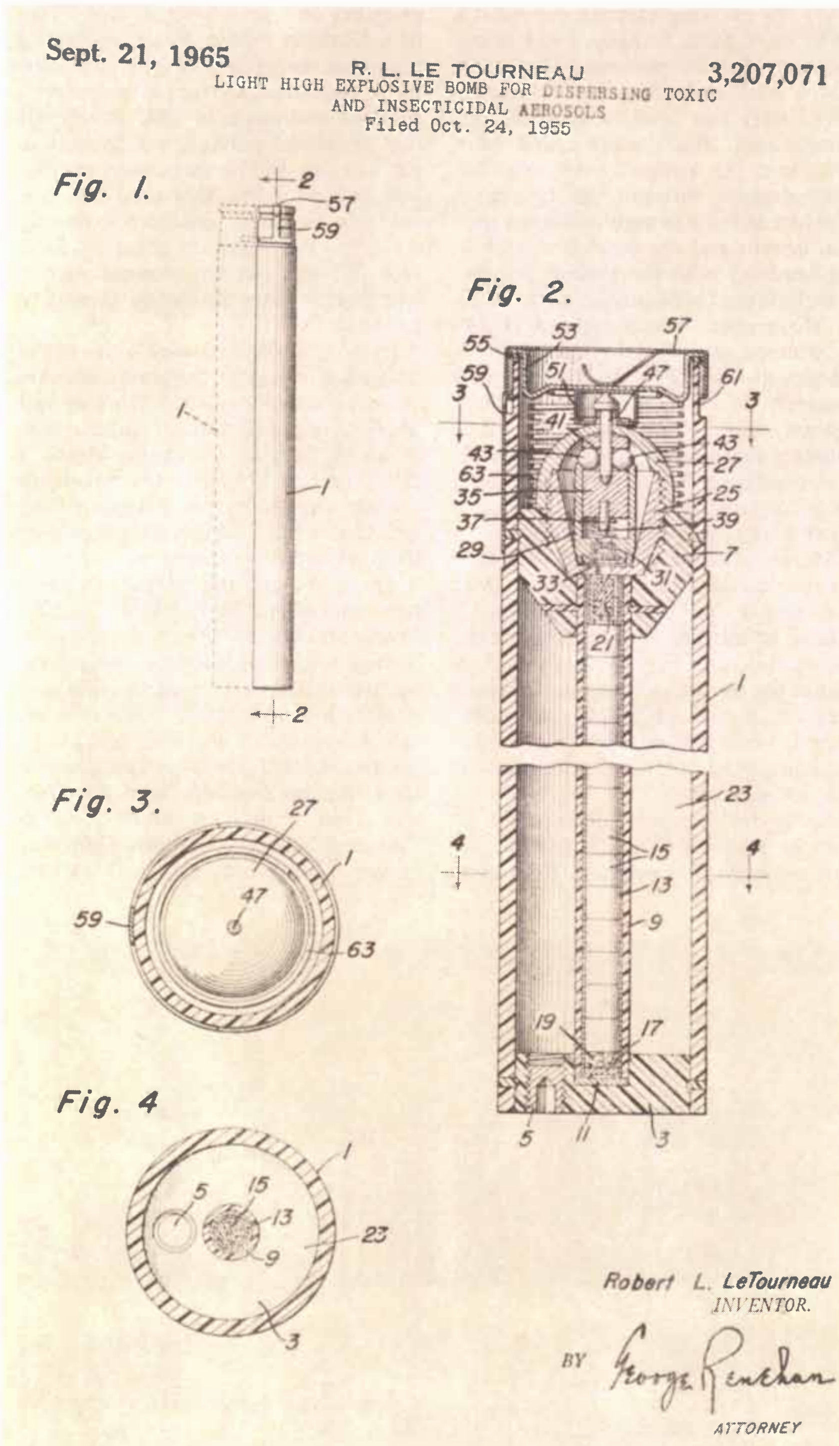
Then, in July, the president's military chief of staff Admiral William D. Leahy and several other advisers conducted in front of Roosevelt what Leahy later called "a spirited discussion of bacteriological warfare." The conversation focused on possible first use to destroy Japan's rice crop. Leahy wrote later that he recoiled from the idea; Roosevelt remained noncommittal. The president never indicated whether he would launch a biological-warfare



"BLACK MARIA," a somber tar-paper building, was constructed in 1944 to house biological-weapons experiments at Camp Detrick. A soldier armed with a submachine gun occupied the wood guard tower at the left. The building was razed soon after the war.

attack in retaliation against Axis first use or whether he would countenance first use against Japan. (At the time claims were circulating that Japan had used biological warfare against China.) In stark contrast to his public pledges that the U.S. would not initiate gas warfare, Roosevelt thus bequeathed to Truman an ambiguous legacy regarding biological weapons.

Two weeks after Truman entered the White House in April of 1945, and a day after the president had received a lengthy briefing on the atomic bomb, Secretary Stimson got a memo from his special assistant Harvey H. Bundy. Bundy wrote that Merck and several other members of the biological-warfare program were proposing the use of chemicals against Japanese food crops. "It is a pretty serious step," the assistant cautioned, "and you may want to speak to the President." Stimson sent a note to Marshall asking to confer with him at his convenience.



PATENT ILLUSTRATION accompanying a 1955 filing by Robert L. Le Tourneau shows an "explosive bomb for dispersing toxic and insecticidal aerosols." The design may be a descendant of those proposed for anthrax bombs during World War II, when Le Tourneau was involved in biological-warfare research. Details of the anthrax bomb are still secret.

From that point until the war's conclusion, emphasis on biological warfare shifted from bacteriological agents to crop defoliants. American scientists certified that the chemicals were not poisonous to humans; the Judge Advocate's Office concluded that their use would be legal because they were nontoxic to people and because the U.S., as a warring nation, "is entitled to deprive the enemy of food and water, and to destroy the sources of supply in his fields."

Stimson, although deeply troubled by the mass killing of noncombatants that American bombing had already caused, seemed prepared to accept the poisoning of Japanese crops. Given that General Marshall wanted to use gas against Japanese troops, he too was probably not unnerved by the tactic of crop poisoning. In May and June an Army Air Force general drew up an elaborate plan for destroying Japan's rice crops by dropping ammonium thiocyanate on rice-producing areas near six major cities: Tokyo, Yokohama, Osaka, Nagoya, Kyoto and Kobe. The commander of the Air Force, Gen. Henry H. Arnold, rejected the plan on tactical rather than moral grounds. Bombing Japan's industry and cities, he judged, would have "earlier and more certain impact."

When other military planners discussed the use of crop poison against Japan, they, like Arnold and his staff, gave primacy to tactical problems. Some questioned whether the supply of chemicals was sufficient; some thought the destruction of the 1945 rice crop would not have any effect until 1946. By then, they believed, the war would have been won and American occupation forces would have the added burden of feeding a hungry civilian population.

On August 3, three days before the bombing of Hiroshima, Arnold's deputy, Lt. Gen. Ira C. Eaker, asked for a comprehensive report on crop destruction by air, including the capabilities of the Air Force, the best chemicals available and the best techniques for their application. He got the report on August 10, the day after the Nagasaki bombing. Four days later the war in the Pacific ended.

ADVISERS FOR BIOLOGICAL WEAPONS gathered at Camp Detrick to consult technical experts. George W. Merck, president of Merck & Co., Inc., is in the middle. Other advisers, from the left, are scientific director Ira L. Baldwin, Capt. Nathaniel S. Prime, Brig. Gen. W. A. Borden, Rear Adm. Julius Zurer, Comdr. William B. Sarles, Colonel Woolpert and Lt. Col. Norman Pyle.

The nation's secretly developed U.S. germ-warfare arsenal was not forgotten in the final months of the war. One high-ranking Army general had commented earlier in the program's history that the Administration might consider a policy of first use against Japan. Later, strategists discussing retaliation concluded that if Japan broke the Geneva Protocol and resorted to gas agents, the U.S. should be prepared to respond with both gas and germ weapons. Adm. Donald B. Duncan, a staff member of the Joint Chiefs of Staff, pointed out that in some situations bacteriological attacks might be more effective than gas.

American beliefs about the morality of biological warfare, however, were never put to the test in World War II. The ultimate decision to use biological weapons would have fallen to Truman; he probably would have relied on the counsel of General Marshall, whom he greatly admired, and of Secretary Stimson, whom he regarded as a moral man. Having sanctioned the use of atomic bombs on Japanese cities, these key advisers might not have taken exception to poisoning rice fields to compel Japan's surrender.

Germ warfare, with its specters of epidemic and invisible poison, might have been harder to endorse. Years later, however, Truman implied in a letter to an associate that if the war in the Pacific had dragged on past mid-August, he would have employed both bacteriological and chemical agents—that, in effect, the atomic bombing he had approved was so much worse.

The U.S. continued the development and stockpiling of biological weapons until 1969, when, in response to the antiwar sentiments of the Vietnam era, President Nixon vowed to halt the programs and destroy the stores. Three years later the U.S. and more than 100 other nations signed the Biological and Toxin Weapons Convention, and many agreed to ban biological warfare outright.

At the beginning of this decade the U.S. began to suspect that the Soviet Union had an active biological-warfare program. The U.S. expanded its own program in response. Suspicions were prompted in part by a 1979 outbreak of anthrax that may or may not have escaped from a facility in Sverdlovsk that the U.S. asserts is a weapons laboratory.

The so-called yellow rain said to have fallen in Laos and Kampuchea has also fueled suspicions. The U.S. Government maintains it is a fungal toxin supplied by the Soviets to Vietnam, although some experts say it is only bee dung [see "Yellow Rain," by Thomas D. Seeley, Joan W. Nowicke, Matthew Meselson, Jeanne Guillemin and Pongthep Akratanakul; SCIENTIFIC AMERICAN, September, 1985].

The clandestine research that began in World War II has thus grown substantially to become an object worthy of international debate. How might peacetime decisions be made? If the past gives any indication, it is probable that the real decision makers are not in the laboratory. In World War II scientists provided the expertise to conceive and develop novel weapons, but historical evidence demonstrates that they lacked any authority to control deployment and use. Both wartime presidents, although decked with the formal authority of the commander-in-chief, knew very little about the biological arsenal over which they presided. Ironically, much of that arsenal had been developed primarily to deal with a country, namely Germany, that never intended to develop a capacity for offensive biological warfare.

American experience during World War II warns that weapons conceived for deterrence or retaliation may become attractive and may seem morally justifiable for offensive strikes. Once the war machine gears up for action, scientists may not be able to constrain use of the technology they have created, particularly in a conflict that is deemed a "just" war.

THE AMATEUR SCIENTIST

Puzzles in two and three dimensions, and ways to simplify their solution

by Jearl Walker

Puzzles that require you to fit pieces into two- or three-dimensional forms are of widespread interest. Can rules be devised to simplify such a puzzle so that you need not consider every possible fitting? To look for rules I began with a common flat puzzle that forms a square. Then I reworked the puzzle on three unconventional game boards that wrap back on themselves. Finally I studied some three-dimensional puzzles invented by Adrian Fisher of Minotaur Designs. They appear to be the most difficult of their kind ever devised.

I began with a flat puzzle consisting of four pieces. How could they fit together to form a square? How many such solutions were there? The puzzle is simple enough so that you may be able to solve it mentally without actually fitting the pieces. I chose to analyze it systematically, hoping to discover rules of procedure that would serve on harder puzzles. My first game board was an empty square with four units on a side. Which piece should be played first onto the board? Since piece *C* extends by only two units vertically and horizontally, it has more possible positions in the square than the larger pieces do; those pieces extend three units in one or more directions. I decided to start with *B*, the largest piece.

B can have four orientations: *T*, inverted *T*, horizontal *T* with its head on the left and horizontal *T* with its head on the right. Since in each orientation the piece can be fitted into the square in four places, the total number of different beginning plays appears to be 16. For this first piece, however, there are really only four beginning positions; the seemingly missing 12 arise because I could rotate the piece in its own plane or turn the board upside down with the piece on it to achieve what at first glance would look like different positions. Such a reduction of orientations can be made only for the first piece played. Thereafter each additional piece must be considered in all its possible orientations. I chose the inverted-*T* orientation for my initial move with *B*.

B fits into the square in four places. Because *B* is symmetric about its vertical axis, however, location *B*2 is reduced to *B*1 if you flip the puzzle. For the same reason *B*4 is the same as *B*3. Therefore *B* has only two unique positions in the square: *B*1 and *B*3. It is not necessary to go further with *B*3 because the "holes" it leaves in units 9 and 13 of the game board are impossible to fill with any of the remaining three pieces.

I next played *A*. Five of the six ways it can be positioned leave impossible holes. The lower illustration on the opposite page shows a promising position (*A*1) and a futile play (*A*2).

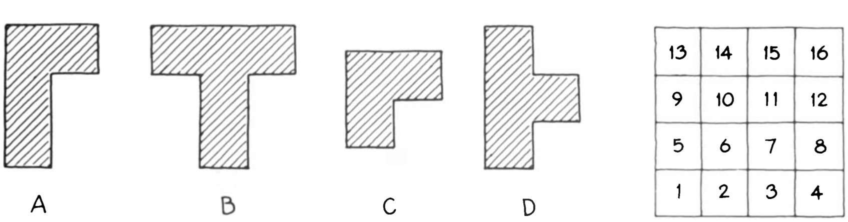
Piece *D* is now added to *A*1. Of the three ways it can be positioned, only one (*D*1) avoids impossible holes. Finally, piece *C* is added to complete the square. Since I have reduced the puzzle to a single route that results in a solution, it has only one unique solution. (Its mirror image is another solution but not a unique one.)

By solving the puzzle in this way I learned several rules of procedure. The first piece played should be large in order to reduce the number of opening positions. The orientation of this first piece can be chosen arbitrarily because any other orientation requires only that the board be rotated in its plane or turned over. If the first piece is symmetric about at least one axis, you need not consider its mirror-image positions.

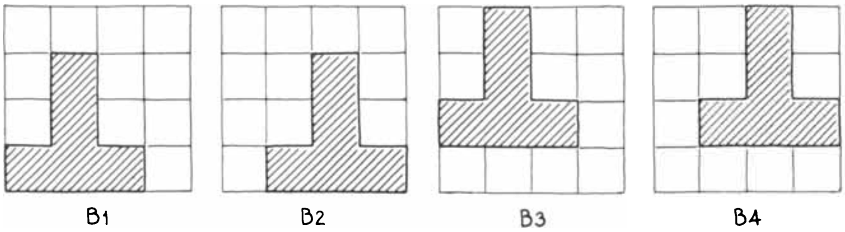
Some positions for a piece can be eliminated if they leave impossible holes, which are spaces that cannot be reached by the remaining pieces or that require unavailable shapes. One or two spaces left isolated along the side of the board are common examples. My game is dull because when the second piece is added, only one of its positions is viable. More interesting puzzles have several viable positions that you must continue to consider as more pieces are added.

While I was studying various puzzles I was also reading *The Shape of Space: How to Visualize Surfaces and Three-dimensional Manifolds*, by Jeffrey R. Weeks. In this fascinating book Weeks describes several games that are played on a flat surface such as my square board. They differ in that the edges of the surfaces are not real because the left and right edges are glued together, as are the top and bottom ones. If you mentally leave the square on the left side, you reenter through the corresponding unit on the right side. Similarly, an exit through the top brings you back into the square through the bottom. The square can be thought of as a flat representation of a torus, or doughnut. Traveling left or right brings you around the full circle of the torus; traveling up or down takes you through the hole of the torus and back to the front. Weeks suggests that games can be played more clearly on this board if identical squares are drawn around the primary square. I call such an arrangement an extended game board.

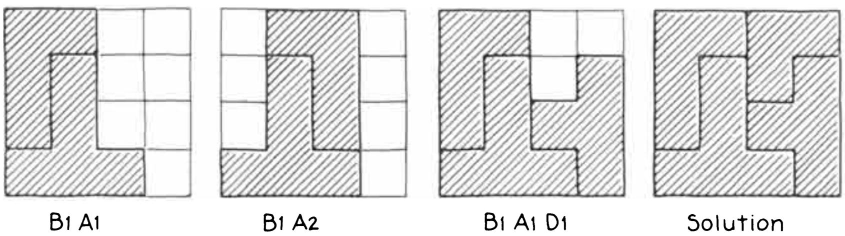
Suppose my four puzzle pieces are played on the new board. Normally a piece cannot be played properly if it sticks out of the square. On the new board such a play is allowed: the part that sticks "out" just sticks back in on the opposite side of the square. An example is shown in the top illustration at the left on page 124. Note that *B*



A four-piece flat puzzle and its game board



Four positions for piece **B**



The four pieces fitted on the board

occupies three separate regions in the primary square.

The puzzle now has many more solutions, but are they unique? Imagine placing the normal solution on the board but shifting it one or more units horizontally, vertically or in both directions. The parts of the pieces that extend off one side of the primary square come back into it on the other side. For example, move the normal solution leftward as a whole by one unit. The leftmost parts of *A* and *B* end up on the right side of the square. Such simple displacements of the normal solutions are not new unique solutions.

A convenient way of representing the normal solution is depicted in the top illustration at the right on the next page. *B* is in its normal orientation and the other pieces are clustered on its right. I call this display an expanded configuration. Imagine placing the configuration anywhere on the new board. As long as the sides of *B* are parallel to the sides of the board, the configuration fills the square neatly. For example, place it so that *B* is in the upper right of the square. Then the other pieces automatically disappear through the top and right sides of the square and reappear through the other sides to fill the square. The mirror image of this solution is also a solution but not a unique one.

Does the puzzle have other unique solutions? Yes, it does; I count four more, each with its own mirror image. One of them is shown in the illustration. Place the solution anywhere on the board; it precisely fills the square. You might enjoy finding the other solutions. You can systematically hunt for them according to the rules of procedure I have given for the normal puzzle. Or you can work with the expanded configurations, mentally flipping portions of the pieces that extend too far to the right or upward. Be wary of counting mirror-image solutions.

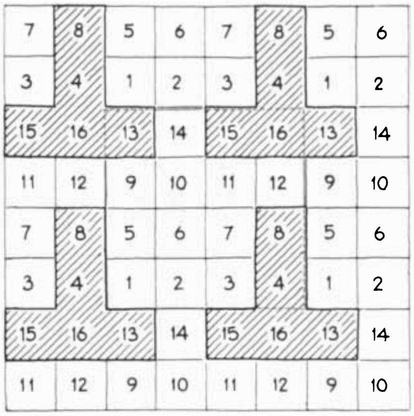
Another way to connect the edges of the game board is with a twist. The top and bottom edges are connected as before, but the left and right edges are mentally twisted and then glued together. This board is a flat representation of a Klein bottle, a topological novelty. If you mentally travel leftward from unit 1, you reenter the square to the left through unit 16. Travel either upward or downward is the same as in the flat-torus board.

The normal solution fits this board but cannot always be displaced successfully. Two more unique solutions are presented in the middle illustration at the right on the next page. Mentally place these solutions on the primary square of an extended board. Note that they can be displaced horizontally but not vertically. Also note that *B* has different orientations in these solutions. Since only the left and right edges are glued together with a twist, different orientations of the first piece played can no longer always be reduced to one orientation by a rotation of the board. How many unique solutions does the puzzle now have?

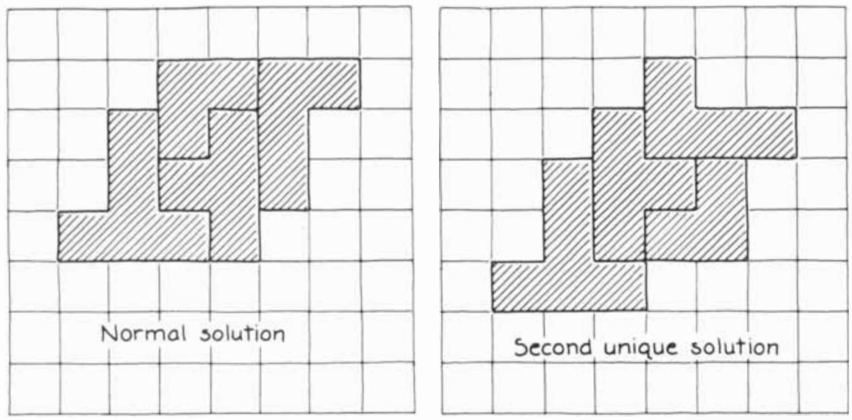
You can make a fourth type of game board by also connecting the top and bottom edges with a twist. (Technically this board is not a flat representation because of the way the corners fit together. Check the lower right-hand corner of the primary square:

only units 4 and 13 glue together there, rather than four different units as in the other glued boards.) Errors are more easily avoided if you play on an extended version of this board. The bottom illustration at the left on this page shows how *B* might be played. It is evident that if the piece is displaced rightward by one unit, the play is no longer valid because the piece then occupies unit 4 twice. How many unique solutions of the puzzle does this game board have?

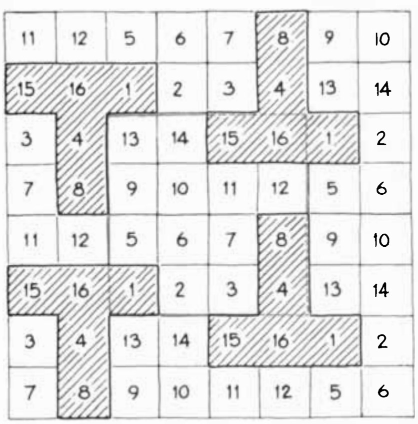
I chose the shapes of the pieces so that a solution on a normal game board is attainable. Is there another choice of shapes for a solution that is possible on only one of the boards? On the flat-torus board five unique solutions are possible with the pieces I played. If the puzzle were changed so that it had only one square piece four units on a side, each of the boards would have only one unique solution.



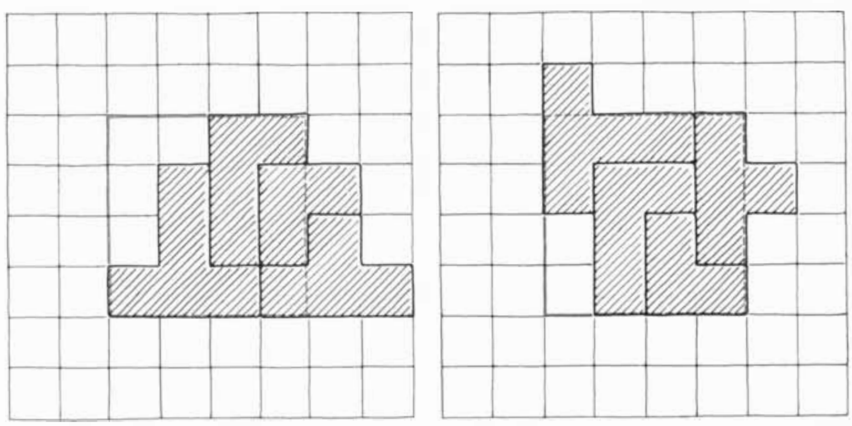
A flat-torus game board



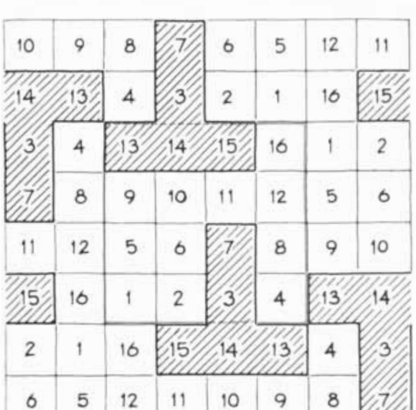
Solutions for the flat-torus board



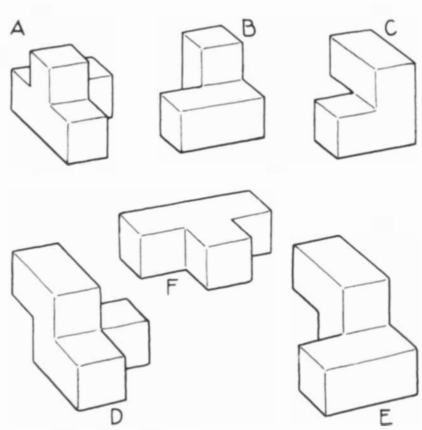
A board with a twisted connection



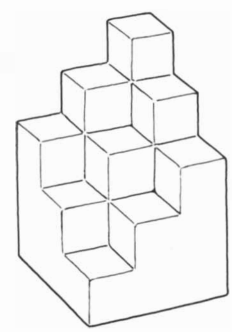
Solutions for the twisted-connection board



A board with two twisted connections



Pieces for the Minotaur Cube



The diamond solution

Imagine other shapes for the pieces, all different. What shapes maximize the number of solutions for each board? You might like to explore larger puzzles to see whether general results can be devised for puzzles of any size.

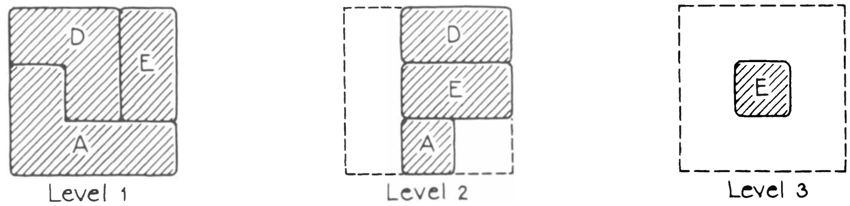
I explored my simple puzzle to warm up for Fisher's three-dimensional puzzles. In developing them Fisher first examined the number of ways basic cubes can be attached to form different pieces, known collectively as polycubes. There are two tricubes (three cubes joined), eight quadcubes, 29 pentacubes and 166 hexacubes.

A popular game of fitting together polycubes to form a cube three units on a side is Soma, invented by Piet Hein and introduced in 1958 [see "Mathematical Games," by Martin Gardner; SCIENTIFIC AMERICAN, September, 1958]. You can find current notes on the game along with a bibliography in Gardner's recent book *Knotted Doughnuts and Other Mathematical Entertainments.* The game employs one tricube and six quadcubes. (The straight tricube and quadcube and the square quadcube are not used.) The game has exactly 240 solutions.

Fisher wondered how the complexity of cube puzzles depends on the number of pieces in it. If the pieces were nothing but 27 individual and unattached cube units or even nine tricubes, the puzzle would hardly be puzzling. It would also be dull if it consisted of only two or three pieces. Fisher concluded that the best puzzle for a cube with three units on a side is one with six pieces. To check this assumption he tested students at a preparatory school in London; the test he set for them was to solve seven cube puzzles whose pieces ranged in number from three to nine.

Recording the average time spent on each puzzle and the percentage of correct solutions, Fisher found that the puzzle with six pieces did indeed take the most time and resulted in the least success. The puzzle with five pieces scored next, closely followed by the one with seven. Fisher designed more complexity into his puzzles by choosing pieces that are approximately the same size but none of which are identical or are mirror images of one another. When they are assembled, none touches a second piece on more than two faces.

One result of Fisher's efforts is the Minotaur Cube, which is made up of three quadcubes and three pentacubes. There is only one solution to form a cube with three units on a side. The Minotaur puzzle can also be assembled into a stepped structure called a diamond solution, which resembles a crystal formation.



Pieces fitted in a cube board

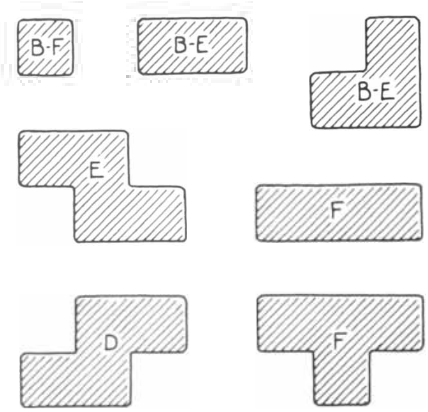
I began my search for the cube solution by listing the possible locations of piece *A,* one of the larger pieces. This time the board was an empty cube. As with the flat puzzles, the orientation of the first piece played is arbitrary because all other orientations are reducible to the chosen one when the board is rotated. For a given orientation *A* has four possible locations. I next added *D,* which has a total of 34 possible positions for all the possible locations of *A.* For each of them I considered all the possible positions of *E.* After three or four pieces had been played I began to eliminate some configurations because they created impossible holes, but still the catalogue was large when I found the solution.

One reason a puzzle with three dimensions is more difficult than a flat one is that impossible holes are not as obvious; a hole left in one level might be filled by a piece played in another level. Impossible holes sometimes show up as an isolated hole or an isolated string of two or three holes on a side of the board. Some holes on the sides or in the interior are impossible in the sense that none of the remaining pieces can fill them.

The upper illustration on this page shows the result of playing the pieces in the sequence *A, D* and *E.* When I played *E,* I was concerned with the hole in unit 12 in the first level because *D* bordered it on two levels. Which pieces can reach down into that hole? *E* can, but playing it makes unit 3 on the second level impossible to fill with any of the remaining pieces. *B* could be played, but it too soon leads to impossibilities.

A more economical way to search for a solution might be to concentrate on the most promising initial fittings. Begin with *A.* How can the other pieces be played so that they occupy part of the remaining empty region in the lowest level? For example, all the pieces can be turned in such a way as to occupy a single unit on that level, but only *F* can be turned so that it occupies three units in a row.

Make a list of these basic shapes.



Basic shapes

Then list the ways they can be combined to fill all the empty region. Four arrangements are shown in the top illustration on the next page. The first two are more promising than the second two because of the number of pieces involved. Remember that the pieces will have to be compatible on the second level. Since the first two arrangements require only two pieces, they should be investigated before the arrangements requiring three pieces.

Which pieces can provide the first arrangement while also being compatible on the second level? There are only two possibilities: either *D* and *B* or *B* and *C.* The rest of the search continues as the empty region on level 2 is filled with the remaining basic shapes. If the search reaches a dead end, return to the second promising arrangement that filled the first level. This procedure might reduce the chore of cataloguing all the possible fittings of the pieces. It does not guarantee success; for that only the full catalogue will do.

The Minotaur Supercube, which has four units on an edge, consists of eight pentacubes and four hexacubes. To say that it is difficult is an understatement. In designing it Fisher determined that the most challenging puzzle of this size is one with 12 pieces. As in the case of the smaller puzzle, the pieces are roughly the same size, but this time no piece is large enough to extend across the entire board.

The Supercube offers several puzzles. There are two ways the pieces can be assembled as a large cube. One of the solutions has a delightful twist. You can pick up several pieces as a whole and reposition them on the remaining part of the cube so that the final assembly is a "diamond solution." Similar stepped structures can also be built employing more than 200 other arrangements of the pieces, but not one of them can be "flipped" into a cube. In addition the pieces can be rearranged to form two cubes (three units on a side) and a small pyramid.

Finding a solution to the large cube formation calls for great patience or keen intuition. To sample the complexity of the puzzle I began with one of the "uglier" hexacubes, turning it so that five of its basic cubes were in the lowest level of the game board. The hexacube had four possible locations in that level. It could be raised to the next level and to the one after that, each time having four more possible locations. Thus the piece has a total of 12 opening plays.

Choosing one of the plays, I added another hexacube. It had 96 possible positions. If two pieces are played, the puzzle has about 1,200 possible routes. I tested one route by playing a third piece, a pentacube. It had 44 possible positions. I explored each of them by considering the holes left in the game board and the remaining pieces. After I had played four or five pieces some impossible holes started to appear. Clearly this puzzle is difficult. Yet I have been told that one person found the solution that flips into a diamond and the solution to the two small cubes and the pyramid in about six hours.

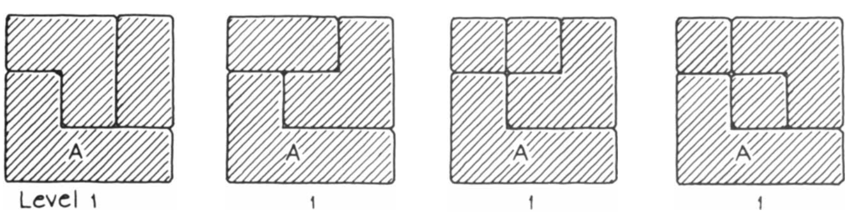
Fisher has also studied larger puzzle structures. Of the 29 different pentacubes, you can form a cube with five units on a side with any 25. There are many such solutions. If you use all the different hexacubes and the square quadcube, you can build a cube that is 10 units on a side.

Fisher has also constructed pyramid puzzles that consist of pieces he calls "wedgehedrons." In designing the puzzle he began with two tetrahedron pyramids that are connected by an octahedron. These units can be connected in three ways so that no two tetrahedron faces touch each other. One of the ways results in a shape Fisher calls a wedge. The pieces for the wedgehedrons puzzle consist of different attachments of two wedges arranged so that no two tetrahedron faces and no two octahedron faces touch each other. When the pieces are assembled, this rule about similar faces not touching also applies to adjacent pieces.

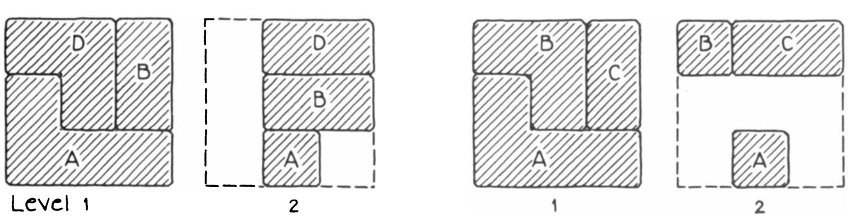
The Minotaur Pyramid consists of five wedgehedrons with four tetrahedrons along each edge. Hidden within it are four empty tetrahedron spaces. The Minotaur Giant Pyramid consists of 10 wedgehedrons, has five tetrahedrons along each edge and contains five empty tetrahedron spaces hidden from view. Five of its pieces are identical with those of the smaller pyramid. Fisher has studied even larger pyramids, some of which require components made from the other two ways of connecting two tetrahedrons by means of an octahedron. All the pyramid puzzles are perplexing because the odd angles and curious "feel" prevent the player from developing any intuition about the fit of the pieces. You might enjoy searching for rules of procedure for solving these puzzles.

I leave you with a final challenge. Can Fisher's smaller cube puzzle be played on a board that has connected sides? The connections can either be direct or be made by means of a twist that is a quarter turn or a half turn. If they are direct, you can assemble the cube in the normal way and then displace it through the sides of the game board. Can you find new unique solutions for the cube on these boards? A simple example is shown in the bottom illustration at the right on this page.

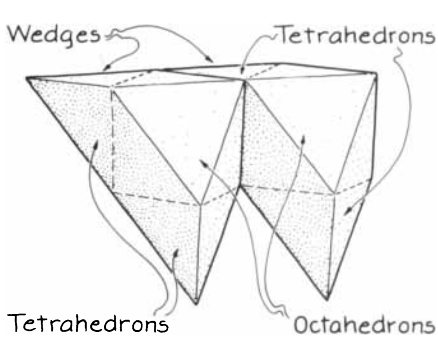
Fisher's cube and pyramid puzzles are available from Minotaur Designs. North American readers should write to 247 Montgomery Street, Jersey City, N.J. 07302; all others should write to 42 Brampton Road, St. Albans, Hertfordshire AL1 4PT, U.K. The smaller puzzles cost $20 (£12.50), the larger ones $38 (£24).



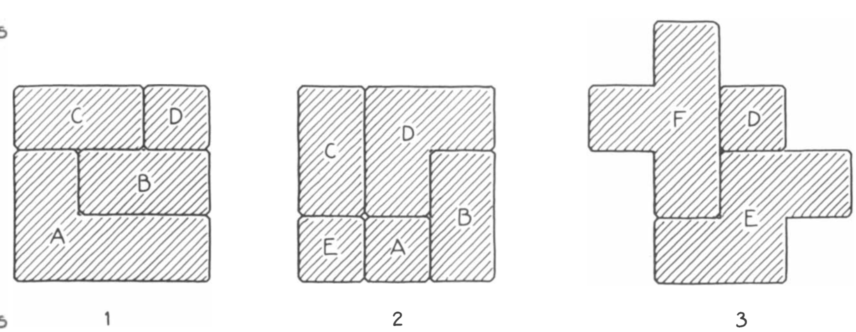
Ways to fill level 1



Compatible pieces



A "wedgehedron" piece



A solution for a cube with glued sides

The Tandy® 6000 HD gives you the competitive edge in XENIX® programming

Streamline your DP operations

Break away from your mainframe with the powerful Tandy 6000 HD personal computer. The 6000 HD offers high speed, large memory capacity, and includes the XENIX multiuser operating system. The high-performance Tandy 6000 HD makes systems development more efficient and cost effective. It's ideally suited for today's software developers.

We support the 6000 HD with our XENIX Development System: an extensive set of programs, utilities and subroutine libraries designed to facilitate "C" language software development.

We also offer popular XENIX-based programming languages such as BASIC, COBOL, FORTRAN and PASCAL. We've got the tools you need to develop the programs that suit your needs.

It pays to share

Since the 6000 HD is a multiuser system, people throughout your office can access it from inexpensive data terminals—for program development, as well as user applications. The Tandy 6000 HD is your key to total office automation.

Real desktop power

The Tandy 6000 HD (26-6022, $3499) features a built-in 15-meg hard disk and 512K internal memory (expandable to 1 megabyte). A unique dual-processor system accepts 16-bit data and processes it as 32-bit "words." And expansion is easy, with four internal slots.

Stop by a Radio Shack Computer Center today . . . we're ready to talk business.

Available at over 1200 Radio Shack Computer Centers and at participating Radio Shack stores and dealers.

TANDY COMPUTERS: In Business . . . for Business™

Radio Shack® COMPUTER CENTERS
A DIVISION OF TANDY CORPORATION

Price applies at Radio Shack Computer Centers and participating stores and dealers. Data terminals, XENIX Development System and programming languages sold separately. XENIX/Reg. TM Microsoft Corp.

COMPUTER RECREATIONS

Algopuzzles: wherein trains of thought follow algorithmic tracks to solutions

by A. K. Dewdney

Problemtown and Solutionville are 100 miles apart. At noon one fine day in June a train leaves each city and heads for the other at 50 miles per hour. Simultaneously a bumblebee that had been resting comfortably on the headlamp of the Problemtown train stirs to life and flies down the track toward Solutionville at 90 miles per hour. When it meets the Solutionville train, the bee makes an abrupt about-face and flies back along the tracks at 90 miles per hour. When it meets the Problemtown train, it does another about-face. It continues reversing until the trains meet. How far does the bumblebee travel?

The answer to this familiar puzzle, which is given toward the end of the department, is short. In fact, it is simply a number. Some puzzles, the ones I call algopuzzles, have more complicated answers. They consist of a recipe, or procedure, for bringing about some desired state of affairs. In other words, the answer is an algorithm. Traditional puzzle literature is full of algopuzzles. We may be asked to divide a container of fluid into three equal amounts by a succession of pouring operations involving containers of specific capacities. We may be asked to row a wolf, a goat and a cabbage across a river by means of successive ferrying operations. We may even be asked to detect a bad penny (overweight or underweight) among many by consecutive weighing operations on a scale.

The difference between the answers to an ordinary puzzle and to an algopuzzle should be made clear by an explicit example. To encourage the proper train of thought, I shall continue along the tracks of the first puzzle. Fortunately the second example does not require the two trains to squash a bumblebee between them in a paroxysm of infinitesimal vibrations. Instead of colliding the trains see each other in time to stop. There is a great rumbling hiss followed by the squeal of steel brakes. (For some reason I picture two leviathans from the age of steam.) Between the trains lies a short stretch of track and a siding that is only long enough to hold one car or an engine [*see upper illustration on opposite page*]. Obviously someone has made a serious scheduling error. The only way out of the dilemma is for the two engineers to use the siding in some manner to get their trains past each other. In this problem and in the problems that follow, each engine and car has a coupler at each end. Each train also has a brakeman who runs along the tracks to couple or uncouple designated cars.

Wiping the sweat from his face with a red polka-dot handkerchief, the engineer from Problemtown declares, "I don't see how we can get by each other. All we got is that little siding over yonder." The engineer from Solutionville is more optimistic. He outlines a plan for getting the two trains past each other. Is it possible? Readers might like to look at the illustration before reading on. To be concrete, the illustration shows two trains with five cars each. In a true algopuzzle, however, trains always have *n* cars.

In arriving at a solution to this rather simple algopuzzle, a typical reader may have mumbled something like the following: "Let's see, now. Why not just put the Problemtown train on the siding, one car at a time. The Solutionville train can then just shuttle back and forth, pulling a new Problemtown car on the siding in one direction, puffing past it through its entire length and then coupling onto it to push it off the siding before puffing back again."

Such a description provides a starting place for an algopuzzle solution, but it must be made clearer and more explicit. For this purpose an algorithmic format is ideal. Some notation will also be helpful. There are four main sections of track that play a role in the solution: the stretch from Problemtown to the siding (designated *A*), the stretch between the siding switches (*B*), the siding itself (*B'*) and the track from the siding to Solutionville (*C*). An instruction such as "Forward *A*" means that the Solutionville train, consisting of the engine and all cars currently coupled to it, moves forward until it lies entirely within the *A* section of track.

Both trains have *n* cars, of course. The engine of the Problemtown train is labeled *P*1 and its cars are labeled *P*2, *P*3 and so on. The instruction "Couple *Pk*" means that the Solutionville train currently has one end adjacent to the *k*th element of the Problemtown train; the Solutionville train rolls gently into *Pk* and couples onto it with the wonderful shuffling noise that those of us who lived near rail yards remember well.

The algorithm can now be given:

```
uncouple P train
for k = 1 to n + 1
   forward to A
   couple Pk
   backward to C
   forward to B'
   uncouple Pk
   backward to C
   forward to A
   backward to B'
   couple Pk
   backward to C
   uncouple Pk
couple P train
```

At first the Problemtown train is completely uncoupled into separate units, one engine and *n* cars. The algorithm then enters a loop in which the same 11 steps are repeated over and over again. The Solutionville train does all the work. Without ever uncoupling any of its own cars, the train proceeds forward to the *A* section of track, where it couples with the first ($k = 1$) unit of the Problemtown train, the engine. It pulls the engine backward to the *C* section and, throwing the siding switch, pushes it onto *B'*, where it uncouples the Problemtown engine. The Solutionville train then backs up to section *C* again, throws the switch and puffs forward into the *A* section. It subsequently backs into the siding, couples once again to the Problemtown engine and pushes it out of the siding to section *C*, where it uncouples the engine. It then repeats the same sequence of steps with each of the Problemtown cars, one at a time. When the main loop is finished, the Problemtown train is on the Solutionville side of the Solutionville train. As soon as it has recovered from the lengthy algorithmic operation it is ready to chug off to Solutionville.

The next example of an algopuzzle comes about as the Problemtown train completes its trip to Solutionville. As he watches the scenery unfold, the engineer realizes with a gasp that he has forgotten his lunch. There is nothing he can do but go back. He applies the brakes and gently eases the massive n-car train to a halt. Backing all the way to Problemtown has little appeal. Fortunately the engineer sees a very short spur line off to his right. It has a switch for either direction of track and is just big enough to hold one car. Inspired by the Solutionville engineer's earlier algorithmic feat, our humble hero is not to be outdone.

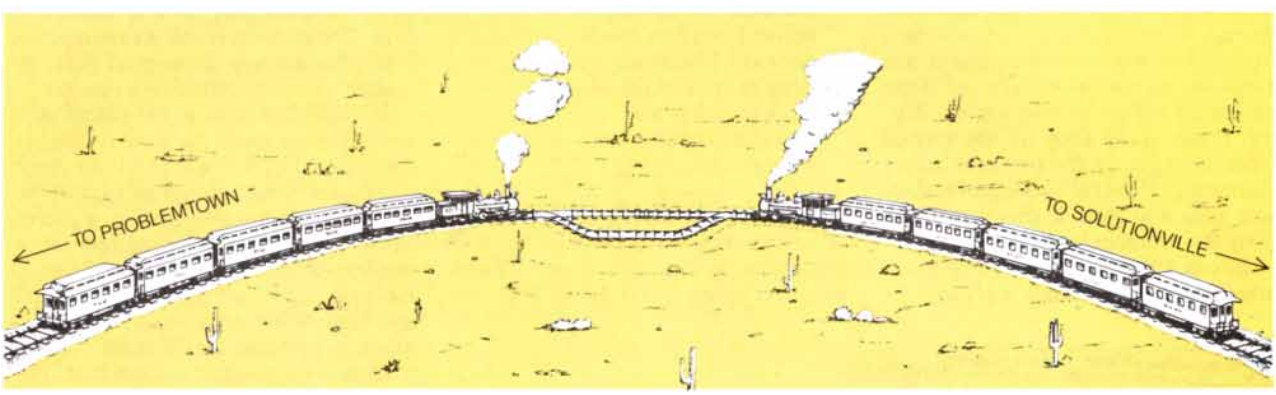
To begin, the Problemtown engineer draws a diagram of the spur on a piece of paper [*see lower illustration on this page*]. Then he writes out an algorithm and checks it carefully by tracing his finger back and forth over the paper and muttering to himself. He then begins a laborious series of uncouplings, shunts and recouplings that eventually results in the reversal of the entire train. Not only is the order of the engine and cars completely reversed but also the engine has been turned around and so has each car. How did he do it? I shall publish the best of the early solutions sent to me. In this case best means both clearest and cleverest. There is more to the problem than meets the eye.

While on the way back to retrieve his lunch, the Problemtown engineer realizes his solution has taken a lot of energy. The coal in the tender is nearly depleted. It turns out that the amount of work done by his train was proportional to n^2: if mass is measured in cars and distance is measured in car lengths, he has moved on the order of n cars a distance of n. Only now, as he chugs home, does he ponder the existence of a more efficient solution. Is it possible to turn the train around in fewer steps, say on the order of $n^{3/2}$ or even $n \cdot \log n$?

For the purpose of training single-track minds further in the gentle art of algopuzzlery, here is a pretty little poser. A lone engine approaches a circular track containing two empty cars. Between the cars is a bridge that is just strong enough to hold one car but not the engine [*see top illustration on next page*]. The engineer must reverse the position of the cars. As in the foregoing algopuzzle, all coupling and uncoupling is done when the train is stationary. (In other words, no "flying shunts" are allowed.) The bridge is no longer than a car, and the engine must be off the circular track when it has finished the job. The solution to this problem is a simple algorithm with no loops. Readers will find the algorithm outlined next month.

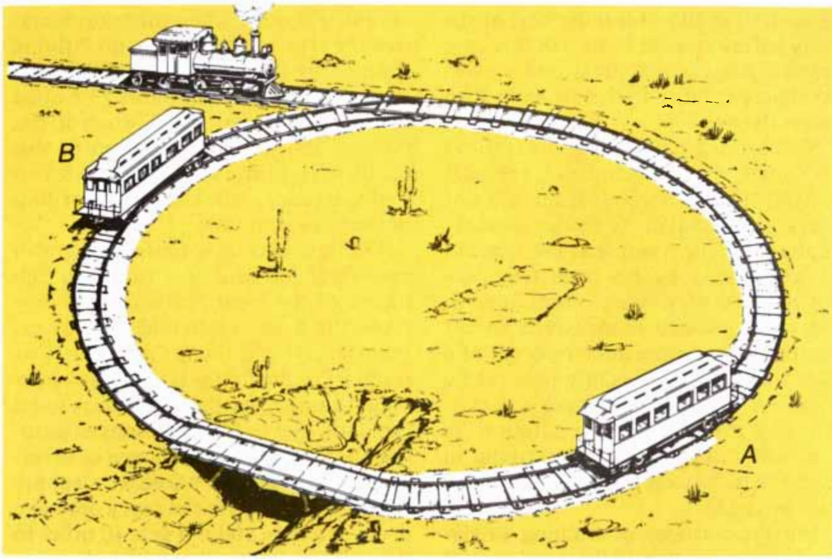
The next two algopuzzles are new to me. They continue the transportation theme of the train problems but they move the inquiry into high gear by using trucks. I call them the Desert Fox puzzles. In the first puzzle a patrol vehicle belonging to the Desert Fox has a fuel tank that holds 10 gallons of gasoline. Its tank is filled from one of several 50-gallon drums stored at a supply depot. If the vehicle can carry only one drum at a time and if it gets 10 miles to the gallon whether it carries a drum or not, how far can it go before running out of gasoline? The answer surely depends on the number of barrels stored at the depot. In the spirit of algopuzzle generality, the number of barrels is specified as n.

If n is equal to 1, the answer is easy. The vehicle fills its tank from the single drum, loads it and sets out into the



How can the trains pass each other?

How does the train turn around to retrieve the engineer's lunch in Problemtown?



The engine must switch the cars without crossing the dangerously weak bridge

hot desert sun on patrol. Refueling itself from the drum as necessary, the vehicle obviously travels 500 miles before it runs out of gas. How far can the vehicle travel on two drums? Since it can carry only one drum at a time, it will carry the first drum some distance out into the desert, refuel from it if necessary, then return to the depot for the second drum. It carries the second drum out on patrol by the same route. It may or may not stop to refuel at the first drum before continuing its journey. It may even drop off the second drum and pick up the first one before continuing. How far can it go by shuttling back and forth in this way? Suddenly the problem looks less trivial.

Here is a short algorithm that guarantees the patrol vehicle a journey of 600 miles. The truck starts its route at the depot:

fuel vehicle and load first drum
forward 100 miles
drop drum and refuel
back to depot for second drum
refuel and load drum
forward 100 miles
refuel from first drum
forward 100 miles
drop drum and refuel
back to first drum
refuel and load drum
forward 100 miles

At this stage in the algorithm the vehicle has progressed a grand total of 200 miles. It has with it two drums. The first drum holds 10 gallons and the second holds 30. The patrol continues when the driver refuels the vehicle from the first drum (leaving it empty), loads the second drum and drives off over the dunes. It therefore travels 400 miles more before running out of fuel. The total distance traveled is 600 miles. Since the improvement in distance over the one-barrel case is only 100 miles, readers would be right to suspect that a better algorithm exists. Indeed, one can achieve what the Desert Fox's enemy might describe as a "beastly number" of miles.

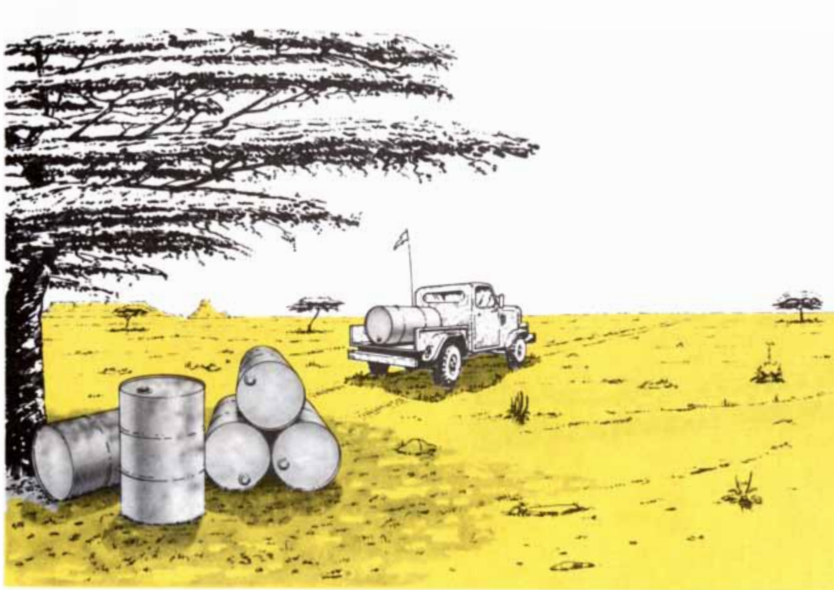
How far can the Desert Fox go if the depot contains n barrels? In the algopuzzle spirit, a single numerical answer or formula simply will not do. One must specify the algorithm that yields the answer.

In another part of the vast sandy domain the Desert Fox (temporarily) calls home, there is an enemy patrol route that is traveled in a very different way: an aircraft drops n drums of fuel at arbitrary points along a closed-loop route. Each drum contains a certain quantity of gasoline but the amount may vary considerably from one drum to the next [*see illustration on opposite page*]. Since the patrol is in unfriendly territory, the car and driver are parachuted in. They land near one of the drums, fuel up and begin the tour. The car's fuel tank is expandable: it can handle any amount of fuel, no matter what size drum it encounters.

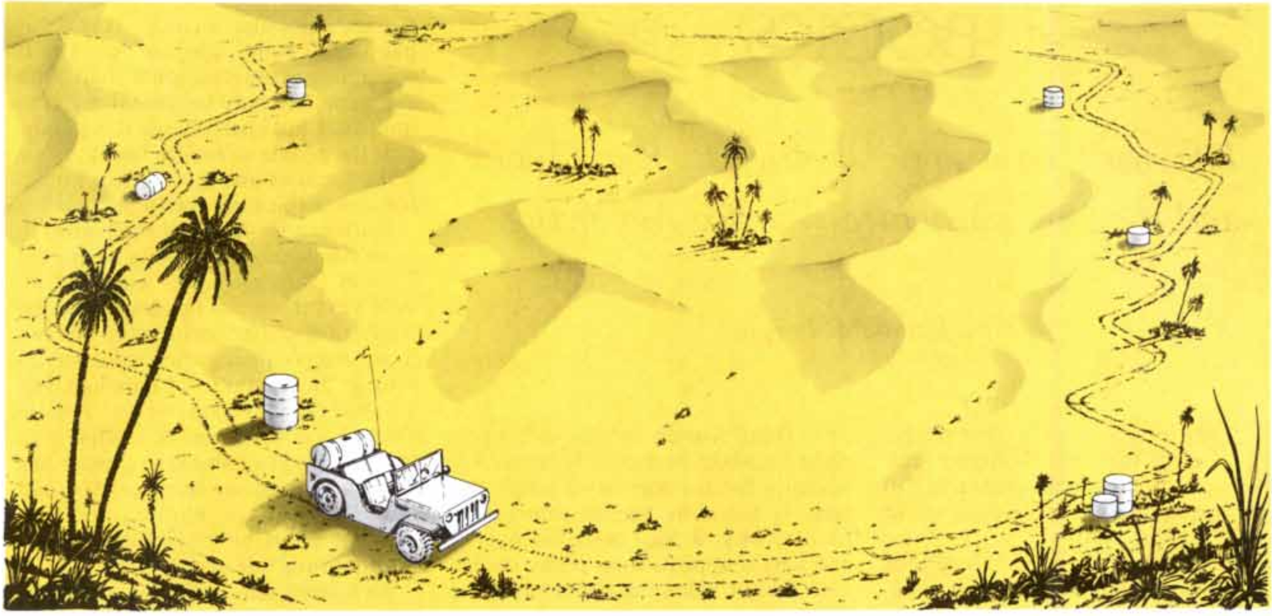
Strangely enough, as the patrol proceeds, the car never runs out of gas before it reaches the next drum. In other words, the total amount of fuel in the drums is exactly enough for a patrol car to complete the loop, neither more nor less. This is odd because, as I said, the placement and quantities of fuel are completely arbitrary. For example, it is perfectly possible for the car to begin its journey at a drum that does not contain enough fuel to get the car to the next drum in either direction. It is even possible to get to the next drum but to be stymied at some point thereafter. How is the feat possible?

The Desert Fox has observed these patrols and is frankly puzzled by them. Is the enemy, known to be mad, also lucky? A subordinate who spends every night in his tent solving puzzles discovers the answer. "It seems, sir, that no matter where the barrels are dropped or how much fuel they contain, there is always a place where the car can begin its patrol with every hope of completing the journey."

I leave it to readers to discover the secret. There is, of course, an algorithm involved. A step-by-step procedure that locates at least one barrel at which the car can begin (and complete) its patrol must first be found.



How far can the patrol truck go on **n** *barrels of fuel?*



Where must the patrol car begin its journey to ensure a successful circuit?

The algorithm must also specify a direction in which the car can travel. To go the wrong way could be disastrous. As with the train algopuzzles, I shall publish the best of the early answers to reach me.

At the beginning of the column I promised to reveal the answer to the puzzle involving the bumblebee and the trains. Good puzzle posers are often careful to mislead readers, particularly when there is an easy answer. By describing the infinite number of reversals in the bumblebee's flight I deliberately tried to confuse readers. It is necessary to realize only that the bumblebee is in flight as long as it takes the trains to meet. They travel half the distance between Problemtown and Solutionville in just one hour. The bee, always flying at 90 miles an hour, therefore travels 90 miles.

There is an apocryphal story about this puzzle that I cannot vouch for. Someone is said to have posed the puzzle to John von Neumann, one of the greatest mathematicians of the century. He is said to have replied instantly, "90 miles, of course."

"Ah," his tormentor said, "I might have known you'd find the easy way of doing it."

"What easy way?" von Neumann responded. He had summed the infinite series in his head.

Reactions to the March column on Valentino Braitenberg's vehicles were enthusiastic and ranged from practical suggestions to philosophical reflections. The vehicles are conceptual: a slab supports a printed circuit, a battery of sensors (usually two eyes) in front and two motors that drive rear wheels. With only the simplest circuits Braitenberg's vehicles are amazingly varied, from lightbulb-loving machines that gaze at luminescent globes in silent adoration to aggressive vehicles that come zooming in from the outer darkness.

Joseph A. Coppola of Sherrill, N.Y., contends that an actual vehicle can be built for roughly $30 with parts such as a slab of plywood and off-the-shelf motors and electronic components. The Coppola vehicle can be conveniently rewired to test different behaviors. Readers interested in such a prospect can write to Coppola at 304 Park Street, Sherrill, N.Y. 13461 for plans, part lists and suppliers. Would it be reasonable to suggest that in one year's time those who have built successful vehicles might send them to the big city for a lark under the streetlights?

Tom Napier of Dresher, Pa., has reminded me of a sophisticated vehicle constructed in the 1950's by W. Grey Walter of the University of Bristol [see "An Imitation of Life," by W. Grey Walter, SCIENTIFIC AMERICAN, May, 1950, and "A Machine That Learns," by W. Grey Walter, SCIENTIFIC AMERICAN, August, 1951]. The vehicle was called a tortoise. According to Napier, it "had a single drive wheel and used two freely rotating fixed-axis support wheels. A steering motor rotated the drive wheel and drive motor about a vertical axis. Both the drive motor and the steering motor turned continuously but were switched to half or full speed by two relays. The relays were operated by two vacuum tubes whose input came from a vacuum photocell that rotated with the steering direction. Since the drive and steering motors had two states but neither ever turned off, the tortoises moved toward light sources in a cycloidal path."

If it is not premature to suggest that someone with a continuing interest in the subject act as a focal point for vehicle communications, I would nominate Gary Blauer, who has volunteered for the job. Blauer would be glad to hear from readers interested in reading (and writing) about vehicles. Any mail should be sent to him in care of Dain Bosworth, Inc., 100 Dain Tower, Minneapolis, Minn. 55402.

To produce vacillating behavior in a Braitenberg vehicle, I proposed employing a circuit consisting of four neurodes: simple, formal computing elements that are the vehicular equivalent of neurons. Frank Palmer of Chicago points out that two neurodes will do just as well as four. In his scheme the wire carrying the input signal is split in two, and each half is connected to a threshold-1 neurode. One of the neurodes inhibits the other. The inhibited neurode transmits what is left of the signal as output.

Robert Baruch of Bayside, N.Y., makes a curiously chilling observation. He agrees with Braitenberg that outside observers will conclude that even simple vehicles display emotions. Baruch speculates that perhaps the sense of our own emotions is derived in the same way. He writes: "It is possible that the human brain, with its myriads of feedback mechanisms, can look on itself as having emotions and a concept known as 'mind,' even though they are nothing but circuits."

BOOKS

Life with Ayla and her friends: Jean Auel and the new phenomenon of Ice Age fiction

by Brian M. Fagan

Until quite recently few people other than archaeologists realized that 20,000 years ago Europe was an arctic landscape, a world very different from our own. If we Holocene human beings thought of Cro-Magnons and Neanderthals at all, most of us regarded them as dry fossils or as brutish figures who lived from hand to mouth by a stone-and-bone technology, occasionally painting pictures of their prey. How then has such an uninviting era of human history, excavated only by archaeologists and cartoonists, suddenly become the object of an intense nostalgia? Blade tools and mammoths have become props in popular novels and great glaciers and wind-blown steppes form the backdrop for epic Ice Age tales. People seem to have found a new Golden Age, an age when *Homo sapiens* first populated the world.

Surely the epicenter of the excitement is Jean M. Auel. The three novels of her Upper Paleolithic cycle "Earth's Children" (*The Clan of the Cave Bear, The Valley of Horses* and *The Mammoth Hunters*) have sold more than two million copies in hardcover alone, according to Crown Publishers, Inc. Paperback press runs have reached some 12 million copies. The three books have been translated into 18 languages and have been published in 22 countries. A film based on *The Clan of the Cave Bear* did disappear from movie theaters rather quickly, but no matter: a fourth novel is being written and two more are projected.

Some 400,000 visitors (presumably many of them Auel fans) made last winter's "Dark Caves, Bright Visions" one of the most successful exhibitions in the 118 years since the doors of the American Museum of Natural History first opened. Now Auel has been joined in the literary marketplace by two respected scholars: the paleontologist Björn Kurtén and the anthropologist Elizabeth Marshall Thomas. Apparently the Ice Age has a secure future. It therefore seems appropriate to ask why it has been discovered and why it appeals to so many people who do not belong to departments of archaeology.

One element is archaeology itself. Workers in the field have in the past 15 years achieved more complete understanding of this early phase of human culture. Careful excavation of thousands of sites on every continent has begun to yield a sophisticated picture of the varied technologies and economic systems and even the forms of social organization through which *Homo sapiens* was able to adapt successfully to the changing rigors of Ice Age environments.

The stone-and-bone paradigm has begun to give way to an awareness of much more complex technologies involving wood and vegetable fiber. New advances in human genetics and studies of ancient dental morphology are beginning to blaze a trail back deep into prehistory, tracing primeval migrations for the first time. Cave paintings, clay sculptures and minute notations on reindeer antler, as well as elaborately and tenderly buried human remains, now provide some tentative insights into Stone Age beliefs. It is clear that this emerging picture of our roots offers human interest in the deepest sense of the term. Kurtén and Thomas are trying to satisfy an inevitable, and profound, curiosity about Stone Age culture as part of our human ancestry.

Every prehistorian has moments of insight when dry artifacts and food remains become a panorama teeming with life, energy and boundless potential. Few of them ever translate these visions into print or write about the Ice Age as something vibrant and alive, a time of people who had feelings and values and an intimate relation with their environment. Kurtén's popular writings reveal his passionate belief that the Ice Age is something scientists should share with everyone, not only because it is fun to look back but also because it is important for people to realize that the past was once alive and that it was a potent force in shaping our own world.

Kurtén, a professor at the University of Helsinki, is a highly respected Finnish paleontologist with a lifetime's experience of the Arctic and the study of Ice Age mammals. He believes science offers a powerful way to look at those who lived long ago. His volume of nonfiction essays, *How to Deep-Freeze a Mammoth,* explored an Ice Age world inhabited by people like ourselves, human beings who had thoughts, feelings, motives and values—not a horrible place full of brutal monsters and vile savages.

As Kurtén sees it, the Upper Paleolithic peoples of Europe and Asia enjoyed a spiritually rich culture reflected in lavishly decorated artifacts and caves. It was, he argues, an age without wars, whose heroes were ordinary people; the culture flourished because of the extraordinary productivity of the mammoth steppe. Kurtén's view of postglacial times is gloomy, however. The climate warmed; the forests spread. The mammoth fauna became extinct and the ranks of other animals were thinned. Man survived "to work in the sweat of his brow, his life span was shortened, cannibalism, slavery, and war became prevalent."

The point Kurtén is making is that science has moved far beyond the stereotypes and assumptions that would have had us believe the life of Ice Age people was merely short, brutal and beastly. One way he attacks such stereotypes is by writing novels. And so we have *Singletusk,* the sequel to an earlier prehistoric chronicle about the Neanderthals—people Kurtén thinks were our own flesh and blood. *Singletusk* distills the critical years during which modern people and Neanderthals lived alongside one another. Tiger, a main character, is one of the first *Homo sapiens* to join a Neanderthal community, where he and his family live peacefully and happily. A tragic accident critically injures his brother. Tiger sends his favorite son, Whitespear, into unknown territory to bring back a powerful healer.

Whitespear's adventures occasionally take on mythic proportions. They highlight a prehistoric world dominated by the Guardians, giant cloud shapes that encourage or threaten and terrify. The events in the book are real: the kinds of encounters, social relations and hazards any Ice Age hunter

---

Brian M. Fagan is professor of anthropology at the University of California at Santa Barbara.

might have experienced. The incidents described, such as a fishing expedition on a lake, are set clearly in season. Sudden storms erupt, or soft early-summer winds blow. Berries ripen; as the people harvest them the reader is captivated by the colors of the fruit and the bountiful joy of an arctic world come alive for a few brief months. When he writes about expert stoneworkers, Kurtén imbues their artifacts with a sensuous and almost lustrous feel; one understands why their owners prized these objects, perhaps as something almost mystical. One is never far from the symbolic world, from the unknown.

Kurtén's vivid and sometimes lyrical prose re-creates the changing seasons. He brings prehistoric hunting alive: "They preferred to come to grips at close quarters, working in pairs, one using a tremendous broad-bladed sticking-spear, the other a heavy hand-axe." His descriptions give a sense of speed and highly developed skill, a rationale for the use of hunting weapons that could only come from hours of studying stone and bone artifacts. Finally, there is the myth of Singletusk, the one-tusked rogue bull mammoth. Whitespear tracks Singletusk for two months. The great mammoth dies in combat with another single-tusked bull in a scene Kurtén swears is based on an actual paleontological find in the New World.

Singletusk is a mystical account of Ice Age life 35,000 years ago. Behind the legends and the descriptions of long vanished environments, artifacts and hunting techniques lies the author's firsthand knowledge of Ice Age paleontology and archaeology. What makes the novel so special is its sheer authority. Yet this scientist is also—rarity of rarities—a master of evocative description and fantasy. Björn Kurtén reminds one of Sibelius. He is attuned to northern environments, ancient and modern.

"My name was Yanan and my story began where it ended, in Graylag's lodge on the highest terrace above the north bank of the Char River." Whereas *Singletusk* transforms the Ice Age into legend, Elizabeth Marshall Thomas' *Reindeer Moon* tells the haunting, sometimes brutal and always realistic story of an ordinary 13-year-old who lived in Siberia about 20,000 years ago. Like Kurtén, Thomas has firsthand experience of hunting and foraging, in her case among the San of the Kalahari. She understands the intricate relations of hunter-gatherers and the natural world around them, both in arctic and tropical latitudes. Instead of writing an adventure story she has chronicled a brief life constrained by forces quite different from anything we know.

Yanan lives with the harsh realities of hunger and cold, of accidental death and fatal childbirth. When her father dies in a hunting accident and her mother dies in childbirth, she and a sibling have to fend for themselves. They travel for many weeks in search of their kin. Then, when they finally reach safety, their return precipitates great social tension (something that happens in San society as well). The tension is caused by the unexpectedness of their return, by uncertainties about their kin affiliations, by their undefined relations with people living nearby—precisely the kind of problems that persist in hunter-gatherer societies to this day.

I have never read a novel that so brilliantly re-creates the constant tension of tribal life: the conversations about kin ties, the flaring quarrels, the ever changing social dynamics that result when people move away to resolve potentially explosive animosities. This is no novel about a Golden Age but a story of a remote, vanished world where symbolic forces are as powerful as the stark facts of human existence in the Arctic. At times Yanan embarks on "spirit journeys" that evoke the lives of animals with innate intimacy. These creatures—the raven, the mammoth, the bear and the cave lion—are the sources of life.

Apparently *Reindeer Moon* began as a nonfiction work but became a fictional reconstruction along the way. The frustration had felicitous results. *Reindeer Moon* has a compelling authenticity that is unique.

Both *Singletusk* and *Reindeer Moon* are fictional works by scientists. The high priestess of Ice Age nostalgia, indeed its probable inventor, is Jean Auel. *The Clan of the Cave Bear, The Valley of Horses* and *The Mammoth Hunters* are clearly mass-market epics that were aimed squarely at a very wide audience. They have succeeded. Auel has taken millions of readers who rarely thought about early human beings unless prompted by viewing the Flintstones back into the Ice Age, to the time when European Neanderthals and *Homo sapiens sapiens* met for the first time.

Auel's enthusiasm for the Ice Age is obviously real; she writes sincerely, touched by her subject. A writer of mere potboilers would hardly have had the stomach to undertake the rigorous scholarship or to withstand the critical professional scrutiny such a novelistic undertaking would invite. Yet Auel is also a professional who set out to achieve a popular effect, and her research is placed squarely in the service of this objective.

Auel's heroine, Ayla, is no ordinary person; she is painted in epic proportions. In *The Mammoth Hunters* Ayla has tamed and rides a horse at a time when the evidence shows human beings had probably not yet domesticated that animal. (Her cavemates are suitably awed.) Her kit contains an extensive pharmacopeia and she is a fairly astute diagnostician. She has also developed a powerful throwing aid for her spear.

Ayla, then, comes across as a kind of Upper Paleolithic Wonder Woman. Clearly "Earth's Children" is popular entertainment (perhaps it is even becoming an industry), and Ayla is a mass-market epic heroine. To fit the mold, to be commercially viable, she must be larger than life, have memorable adventures, be a trend setter. That her accomplishments are equal to those of any male even lends a flavor of woman's liberation to the story. Of course she must also be sexually active, sensitive and emotionally aware. Her character can be found, by and large, in many popular novels. Ayla is an earnestly drawn formula heroine placed in an Ice Age setting. And her very contemporariness helps to make the reader at home there too.

How does science fare? Some of Auel's descriptions of mammoth hunters' dwellings are impeccable; they were clearly based on lengthy perusal of the academic literature and on interviews with Olga Soffer of the University of Illinois at Urbana-Champaign and other prehistorians who have firsthand knowledge of sites, dwellings and artifacts. Auel used an obscure and remarkable Soviet monograph on mammoth-bone musical instruments to add vivid detail to a shaman's chant. In *The Valley of Horses* she gives a fine description of just how diverse many Upper Paleolithic hunting and foraging cultures in Ice Age Europe must have been. Some parts of the books are altogether engrossing; they bring Upper Paleolithic hunting and technology to life.

Yet, perhaps for the reader's comfort, Auel has sanitized the Ice Age milieu. No matter how skillfully they were crafted, human habitations must have offered those who lived in them a variety of intense odors. We might consider this appeal to a neglected sensorium unpleasant. Our forebears probably did not. The smells of their domestic environment may have signaled who was present and who was absent, the state of health of the group and the nature of the food supply. At times Auel's Ice Age dwellings seem to be as squeaky clean as suburban ranch houses.

No one lives forever; the reality behind that truism, in an Ice Age band, was that a typical human life spanned perhaps 25 years and women probably perished in childbirth at a high rate. Yet Auel minimizes or ignores such facts of life.

It is clear that Ayla is beautiful in a modern sense. Poetic license allows a novelist to stray from reality, but surely a contemporary admirer (Ayla has two of them in *The Mammoth Hunters*) would perceive and respond, however positively, to the presence of axillary hair and other aspects of Ice Age toilette. Her romances—both their emotional and physical expression—could be transposed to almost any other historical setting virtually unchanged—except for one detail: in her bag of botanicals Ayla carries a contraceptive agent. This invocation of Ice Age knowledge of at least some medicinally effective plants enables Auel to endow her heroine with a capacity for carefree adventure that would be the envy of any Club Med.

This is Romance, Adventure, Escape in the tradition of Edgar Rice Burroughs, Cecil B. DeMille or Margaret Mitchell. The Ice Age, then, has become the newest unexplored fantasy frontier back to which a contemporary *Homo sapiens* can flee when the world we have constructed for ourselves becomes intolerable. Auel has charted new territory for those who have become jaded by the Old West, the Old South, the Middle Ages, Outer Space and perhaps even National League football.

Few scientists can cavil at the way Auel has tried to marshal accurate data on Upper Paleolithic lifeways, but they will probably be less accepting of her efforts to reconstruct Stone Age beliefs and social life. The basics are there—the notions of cooperation in the chase, of trading between neighboring communities, of binding kinship ties and of marriage outside the band. But I found that the way in which the characters showed feelings, interacted with one another and related to the symbolic world rang false. One is always aware that Ayla's speech, her emotions and her relationships are familiar, indeed are indistinguishable from contemporary norms. She is a heroine who just happens to be wearing Ice Age garments and handling Ice Age artifacts. We can be sure ordinary Upper Paleolithic people reacted quite differently.

This is where an anthropologist such as Thomas, who has actually lived with hunter-gatherers for a long time, has a literary advantage over Auel. It is for this reason that *Reindeer Moon* is a much more convincing portrayal of what it must have been like to be an Upper Paleolithic woman. Thomas' novel is about average people living in a highly symbolic world, one where the tensions and stresses of band life explode and subside, where people can spend hours discussing social relations, gossiping and arguing; their society is very different from the well-ordered Ice Age theme park graced by Ayla and her pet horse. In *Reindeer Moon* no one is larger than life, except the hidden forces that control the world. In "Earth's Children" the characters assume epic proportions.

One need not fault Auel for this: her genre is quite different. Her novels are a popular epic, a tour through Upper Paleolithic life. Kurtén's and Thomas' novels give them, as scientists, a vehicle for intelligent speculation based on science. Their novels are "walkabouts" through a brief period of Ice Age life, books about the changing verities of the season, about what were at the time ordinary people. It is the imprint of authenticity, mingled with an appreciation of how such natural phenomena as clouds give nature a symbolic form, that makes the Kurtén and Thomas works distinctive. These books will never spawn an informal cult, but they paint a compelling picture of a time of extraordinary human achievement.

Many of the current debates about *Homo sapiens* are buried in the scientific literature; most of them will never reach the ear of the general public. If there were more writers like Kurtén and Thomas, capable of combining science and fiction, we might see an entirely new literary genre: archaeological fiction—a way of bringing current research into the prehistoric past to a much wider audience. By weaving a compelling story around scrupulously accurate scientific data and imaginative reconstruction, the archaeological novelist could bring the past to life, free of the deadening constraints of the academic forum. At the same time current research, and the controversies that surround it, would reach a much broader audience. Until such a literary genre emerges, mass-market audiences will have a less than rigorous picture of what the Ice Age was like.

The appeal of Ice Age fiction may reflect the difficulty lay people have in getting a sense of the Upper Paleolithic by visiting museums or even painted caves in the Dordogne: the paintings are diffused by bright electric lights and the artifacts are dispersed through dozens of museums all over the world.

Now and then a special exhibition does provide a unique impression of the remote past. *Dark Caves, Bright Visions: Life in Ice Age Europe* is the catalogue of the highly successful one at the American Museum of Natural History that brought scattered masterpieces of Upper Paleolithic *art mobilier* together under one roof. Randall White's lavishly illustrated guide is remarkable for its clear exposition and thoroughly up-to-date approach. He points out that the major outlines of the period are now known. The future holds less promise of spectacular discoveries in the ground. Rather, months of tedious analysis in the laboratory await the Stone Age archaeologist. Research centers not on when and where but on how and why.

White's cautious description of Upper Paleolithic life contrasts with that of the novelists. Ethnographic analogy, for example, now appears to be less helpful than scholars of a generation ago would have had us believe. As the late André Leroi-Gourhan warned, we should never let the Upper Paleolithic speak to us "in the accents of 19th-century Tierra del Fuego or the contemporary Sudan."

White believes band size in the Upper Paleolithic fluctuated widely from season to season, partly because of scarce food resources but also as a way of resolving conflicts. He argues that much Upper Paleolithic artistic activity took place when the bands came together to trade and for tribal ceremonies such as initiations in dark caves.

He is cautious too about social organization and ritual. He does hint that differences between burials at the same site might reflect differences in social status and that some findings at ritual caves such as Tuc d'Audoubert in the Pyrenees suggest a link between age and status. Like his scientific predecessors, he puzzles over the art. Did the Magdalenians track time with lunar calendars? Was the cave art a way of bringing animals alive or a storehouse of environmental knowledge passed on from generation to generation? We simply do not know.

Thousands of visitors saw the "Dark Caves, Bright Visions" exhibits. They peered at finely engraved bison, Venus figurines, even a superb reconstruction of a Mezhirich mammoth-bone house. Many of them had read Jean Auel; indeed, I heard several of them talking about Ayla. It was as if they were feeling at least some affinity with the artists and hunters, a sense that the intellectual processes and emotions of individuals who lived 15,000 years ago were somewhat akin to those of people like ourselves. Their reactions made me wonder if Auel had rekindled a longing for simpler and therefore happier times.

In one way this wave of nostalgia is timely. In 1984 the American Museum mounted "Ancestors," a display that highlighted fossils of apelike people far removed from the talented artists and hunters of "Dark Caves." The exhibition came at a time when paleoanthropologists were moving away from images of "man the hunter," of primeval hominids who were expert foragers and big-game hunters. Now the peering eye of the archaeologist's microscope has replaced hunting with scavenging, and it descries a much less flattering picture of early human achievement than was seen in the halcyon days of the early 1970's. Our earliest ancestors seem more apelike. It is hard for even sophisticated late-20th-century adults to feel much affinity with apelike hominids.

"Dark Caves, Bright Visions" was something rather different. It commemorated a brilliant creative explosion, a catalytic moment in human prehistory, involving people like ourselves. It is probably fair to regard that moment as the emergence of human culture: a conscious tradition of learning, of social relations and of the all-important connection between society and the natural world. Such a realization surely is the most fundamental reason for the sudden popularity of and interest in the Ice Age.

As White points out, the archaeologist now stands at the frontiers of what the spade and the microscope can tell about a remote past. The material remains of the Upper Paleolithic are the stuff of which both scientific and fictional visions can be made. The visions are elusive, and that is why respected scientists such as Kurtén and Thomas have turned to the novel as a way of painting the wider picture, of moving into pastures forbidden to scholarly publication.

A popular writer such as Auel exploits many of the same materials to create a quite different fictional vision. Yet thanks to her, an enormous audience has come to realize that as the last glaciation receded it left human culture on its moraines and steppes. Because of these books, everyone should ultimately learn a great deal more about what may well have been a Golden Age after all.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

DROUGHT IN AFRICA

POLITICAL AND ECONOMIC ORIGINS OF AFRICAN HUNGER. Michael F. Lofchie in *The Journal of Modern African Studies,* Vol. 13, No. 4, pages 551–567; December, 1975.

POVERTY AND FAMINES: AN ESSAY ON ENTITLEMENT AND DEPRIVATION. Amartya Sen. Oxford University Press, 1981.

DROUGHT AND HUNGER IN AFRICA: DENYING FAMINE A FUTURE. Edited by Michael H. Glantz. Cambridge University Press, 1987.

MOLECULAR MECHANISMS OF PHOTOSYNTHESIS

THE PHOTOSYNTHETIC BACTERIA. Edited by Roderick K. Clayton and William R. Sistrom. Plenum Press, 1978.

PHOTOSYNTHESIS, VOL. 1: ENERGY CONVERSION BY PLANTS AND BACTERIA. Edited by Govindjee. Academic Press, 1982.

MOLECULAR GENETICS AND THE LIGHT REACTIONS OF PHOTOSYNTHESIS. Douglas C. Youvan and Barry L. Marrs in *Cell,* Vol. 39, No. 1, pages 1–3; November, 1984.

THE STRUCTURAL BASIS OF PHOTOSYNTHETIC LIGHT REACTIONS IN BACTERIA. Johann Deisenhofer, Hartmut Michel and Robert Huber in *Trends in Biochemical Sciences,* Vol. 10, No. 6, pages 243–248; June, 1985.

GRAVITATIONAL WAVE OBSERVATORIES

GRAVITATION. Charles W. Misner, Kip S. Thorne and John Archibald Wheeler. W. H. Freeman and Company, 1973.

SOURCES OF GRAVITATIONAL RADIATION. Edited by Larry L. Smarr. Cambridge University Press, 1979.

THE SEARCH FOR GRAVITY WAVES. P. C. W. Davies. Cambridge University Press, 1980.

THE ANATOMY OF MEMORY

A MEMORY SYSTEM IN THE MONKEY. M. Mishkin in *Philosophical Transactions of the Royal Society of London,* Series B, Vol. 298, No. 1089, pages 85–95; June 25, 1982.

MEMORIES AND HABITS: TWO NEURAL SYSTEMS. Mortimer Mishkin, Barbara Malamut and Jocelyne Bachevalier in *Neurobiology of Learning and Memory,* edited by Gary Lynch, James L. McGaugh and Norman M. Weinberger. The Guilford Press, 1984.

THE AMYGDALA: SENSORY GATEWAY TO THE EMOTIONS. John P. Aggleton and Mortimer Mishkin in *Emotion: Theory, Research, and Experience, Vol. 3,* edited by Robert Plutchik and Henry Kellerman. Academic Press, 1985.

AMYGDALECTOMY IMPAIRS CROSSMODAL ASSOCIATION IN MONKEYS. Elisabeth A. Murray and Mortimer Mishkin in *Science,* Vol. 228, No. 4699, pages 604–606; May 3, 1985.

COLLAPSING VOLCANOES

CATASTROPHIC DEBRIS STREAMS (STURZSTROMS) GENERATED BY ROCKFALLS. Kenneth J. Hsü in *Geological Society of America Bulletin,* Vol. 86, No. 1, pages 129–140; January, 1975.

VOLCANIC DRY AVALANCHE DEPOSITS--IDENTIFICATION AND COMPARISON WITH NONVOLCANIC DEBRIS STREAM DEPOSITS. Tadahide Ui in *Journal of Volcanology and Geothermal Research,* Vol. 18, Nos. 1/4, pages 135–150; October, 1983.

LARGE VOLCANIC DEBRIS AVALANCHES: CHARACTERISTICS OF SOURCE AREAS, DEPOSITS, AND ASSOCIATED ERUPTIONS. Lee Siebert in *Journal of Volcanology and Geothermal Research,* Vol. 22, No. 3/4, pages 163–197; October, 1984.

CATASTROPHIC DEBRIS AVALANCHE DEPOSIT OF SOCOMPA VOLCANO, NORTHERN CHILE. P. W. Francis, M. Gardeweg, F. Ramirez and D. A. Rothery in *Geology,* Vol. 13, No. 9, pages 600–603; September, 1985.

DIVING ADAPTATIONS OF THE WEDDELL SEAL

WEDDELL SEAL: CONSUMMATE DIVER. Gerald L. Kooyman. Cambridge University Press, 1981.

SEAL LUNGS COLLAPSE DURING FREE DIVING: EVIDENCE FROM ARTERIAL NITROGEN TENSIONS. Konrad J. Falke, Roger D. Hill, Jesper Qvist, Robert C. Schneider, Michael Guppy, Graham C. Liggins, Peter W. Hochachka, Richard E. Elliott and Warren M. Zapol in *Science,* Vol. 229, No. 4713, pages 556–558; August 9, 1985.

HEMOGLOBIN CONCENTRATIONS AND BLOOD GAS TENSIONS OF FREE-DIVING WEDDELL SEALS. Jesper Qvist, Roger D. Hill, Robert C. Schneider, Konrad J. Falke, Graham C. Liggins, Michael Guppy, Richard E. Elliott, Peter W. Hochachka and Warren M. Zapol in *Journal of Applied Physiology,* Vol. 61, No. 4, pages 1560–1569; October, 1986.

MICROCOMPUTER MONITOR AND BLOOD SAMPLER FOR FREE-DIVING WEDDELL SEALS. Roger D. Hill in *Journal of Applied Physiology,* Vol. 61, No. 4, pages 1570–1576; October, 1986.

THE CONNECTION MACHINE

THE CONNECTION MACHINE. W. Daniel Hillis. The MIT Press, 1985.

MASSIVELY PARALLEL COMPUTERS: THE CONNECTION MACHINE AND NON-VON. Richard P. Gabriel in *Science,* Vol. 231, No. 4741, pages 975–978; February 28, 1986.

SPECIAL ISSUE ON PARALLELISM. *Communications of the ACM,* Vol. 29, No. 12; December, 1986.

THE BIRTH OF THE U.S. BIOLOGICAL-WARFARE PROGRAM

THE CHEMICAL WARFARE SERVICE: FROM LABORATORY TO FIELD. Leo P. Brophy, Wyndham D. Miles and Rexmond C. Cochrane. U.S. Government Printing Office, 1959.

CHURCHILL'S SECRET BIOLOGICAL WEAPONS. Barton J. Bernstein in *Bulletin of the Atomic Scientists,* Vol. 43, No. 1, pages 46–50; January/February, 1987.

NEW DESIGNS FOR BIOLOGICAL WEAPONS. Susan Wright in *Bulletin of the Atomic Scientists,* Vol. 43, No. 1, pages 43–46; January/February, 1987.

SVERDLOVSK AND YELLOW RAIN: TWO CASES OF SOVIET NONCOMPLIANCE. Elisa D. Harris in *International Security,* Vol. 11, No. 4, pages 41–95; Spring, 1987.

THE AMATEUR SCIENTIST

THE SHAPE OF SPACE: HOW TO VISUALIZE SURFACES AND THREE-DIMENSIONAL MANIFOLDS. Jeffrey R. Weeks. Marcel Dekker, Inc., 1985.

KNOTTED DOUGHNUTS AND OTHER MATHEMATICAL AMUSEMENTS. Martin Gardner. W. H. Freeman and Company, 1986.

COMPUTER RECREATIONS

AMUSEMENTS IN MATHEMATICS. Henry Ernest Dudeney. Dover Publications, Inc., 1958.

How to process documents instead of words.

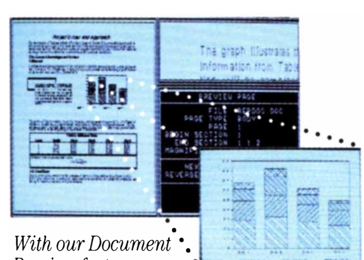
Ho hum, your word processor enables you to process words. But today, when you're creating long, complex documents, you often have to process much more. Graphics, equations, columns, tables, etc.

Lotus Manuscript™ is a complete document creation system that can handle documents from 1 to 800 pages. Manuscript can mix text and graphics on the same page, graphics from Lotus® 1-2-3®, Symphony® and Freelance® Plus. It can also import spreadsheets and charts, plus diagrams and scanned images.

With our Integrated Outliner you can collapse a document to navigate the outline, or move any size section with just a few keystrokes.

You can globally format an entire Manuscript document, or format by sections.

Manuscript's Document Compare feature highlights changes between revisions for foolproof proof-reading.



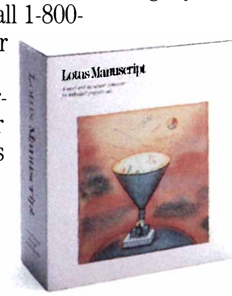
With our Document Preview feature you can see how any page will look before it's printed. Zoom capability lets you take a closer look at graphics and equations.

Also, Manuscript automatically sizes and generates math equations as well as Greek symbols, diacritical marks and brackets.

Our powerful Print Formatter gives you control over the look of your document, from position and size of graphics, to fonts, point sizes and more.

Manuscript also takes full advantage of today's printing technology, from dot-matrix to laser, including PostScript® printers.

Manuscript is designed to work on most IBM® PCs and compatibles.* Its familiar 1-2-3 interface makes it easy to use. And our Manuscript evaluation kit makes it easy to try. For $10.00, you'll get a presentation disk, working software, and a tutorial manual. To get your evaluation kit, call 1-800-345-1043, ask for lot #BD-1450. Or, for more information, see your authorized Lotus Dealer, or write Lotus Development Corp., 90 Annex, Atlanta, GA 30390-0370.



Lotus Manuscript™