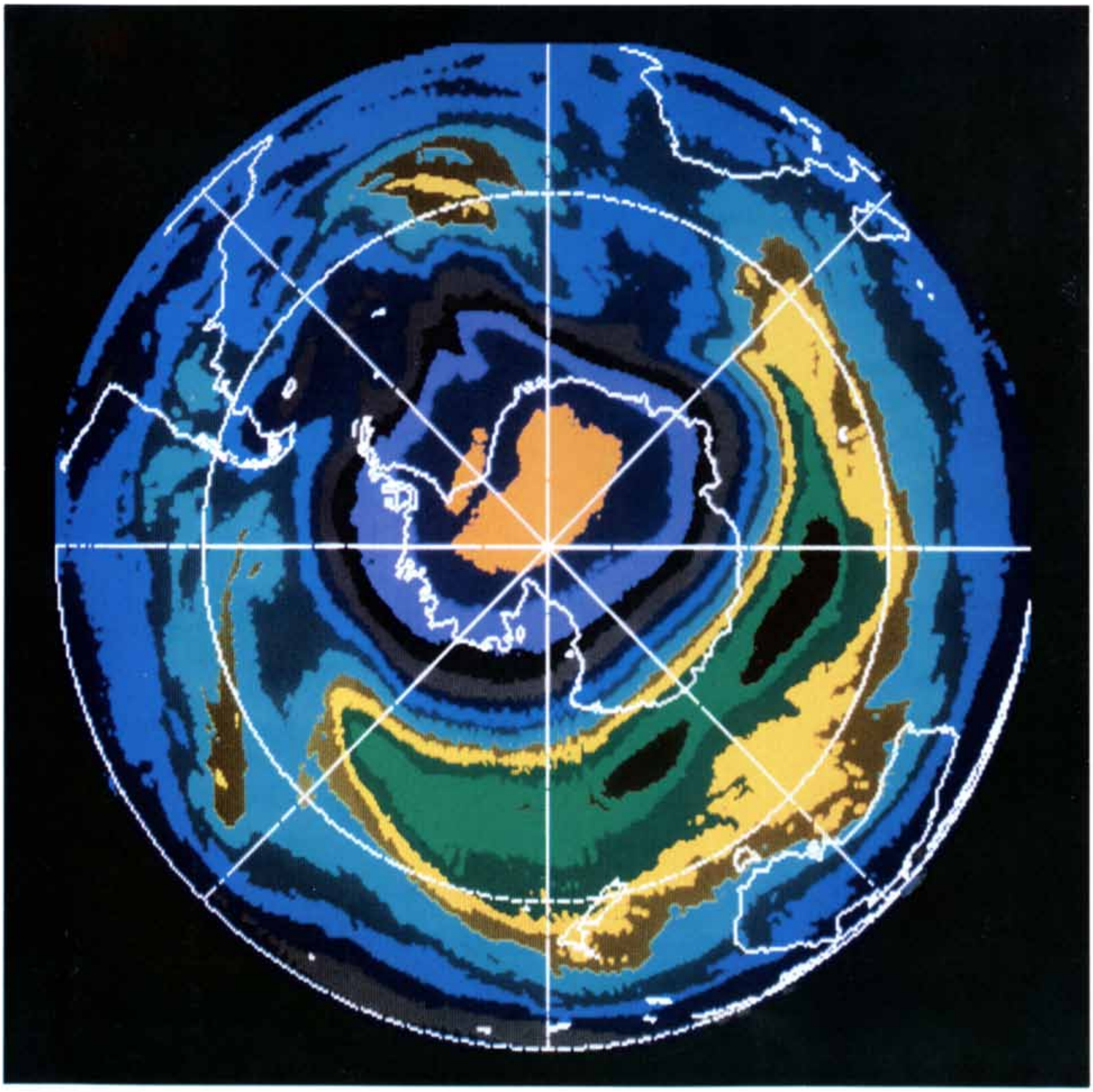
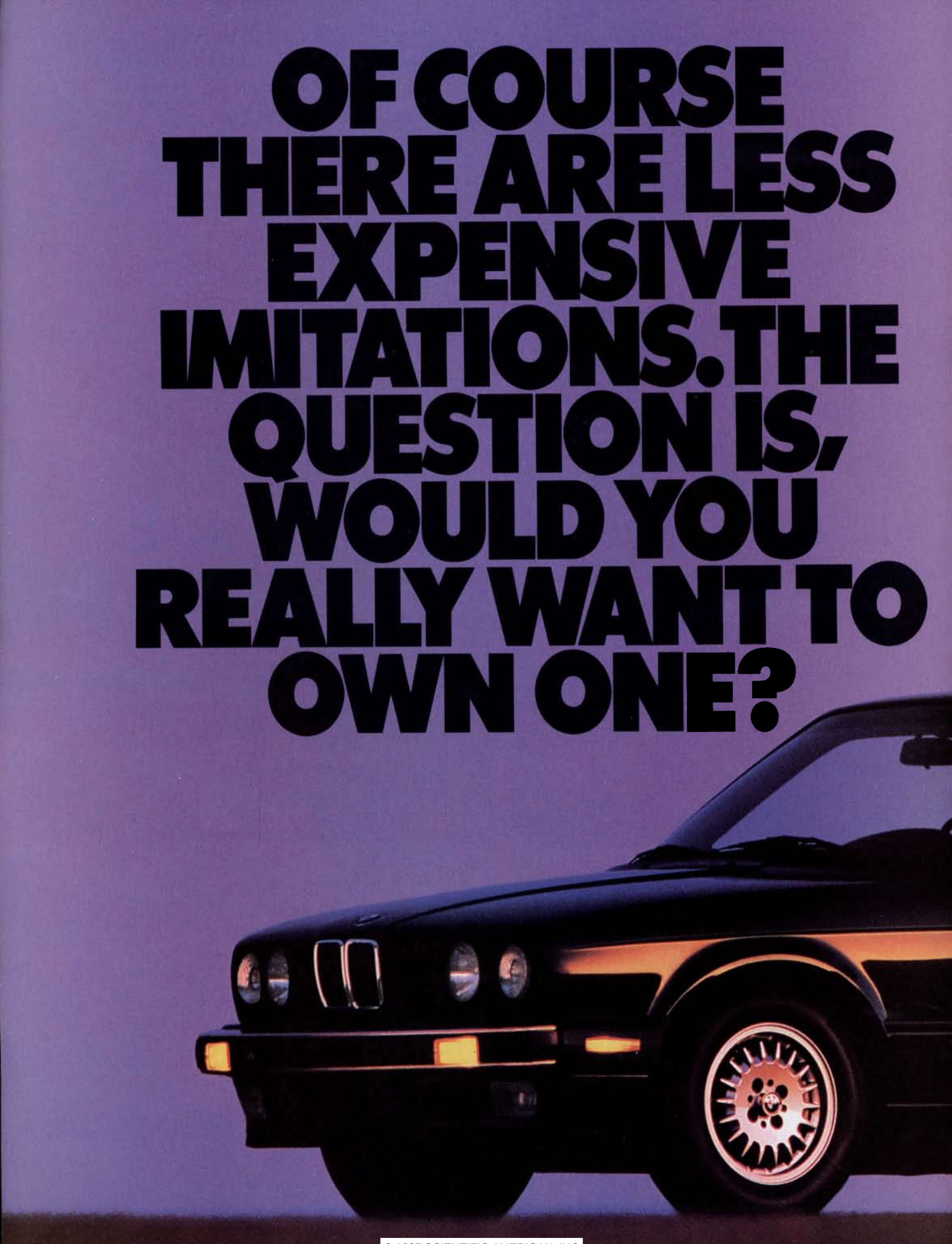
SCIENTIFIC AMERICAN

JANUARY 1988
$2.50

Killer cells fight disease by drilling holes in their targets.
Coming soon: a radio telescope 5,000 miles wide to probe the cosmos.
When the tide ebbs, some fishes crawl and breathe air to survive.



***Ozone Hole** over Antarctica. Does it mean the stratospheric layer that screens out dangerous solar ultraviolet rays is in jeopardy?*


OF COURSE THERE ARE LESS EXPENSIVE IMITATIONS. THE QUESTION IS, WOULD YOU REALLY WANT TO OWN ONE?

SCIENTIFIC AMERICAN

January 1988 Volume 258 Number 1

30 

The Antarctic Ozone Hole
Richard S. Stolarski

A thin layer of ozone in the stratosphere protects human beings and other life forms by absorbing most of the ultraviolet radiation of the sun. The discovery of a springtime ozone "hole" over the South Pole suggests that the entire ozone layer may be thinning. What causes the hole—a global rise in pollutants or a natural change in Antarctic air movements?

38 

How Killer Cells Kill
John Ding-E Young and Zanvil A. Cohn

Killer lymphocytes, the commandos of the immune system, attack tumor cells and cells infected by viruses. They kill by secreting protein molecules that link to form pores in target cells; the cells promptly leak to death. Study of the process may make it possible to improve the killers' efficiency in fighting cancer and such viral infections as AIDS.

46 

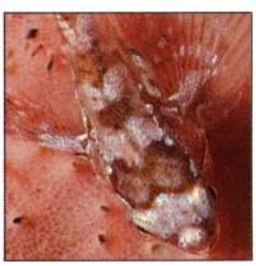
The Reality of the Quantum World
Abner Shimony

Is a photon a particle or a wave? That depends, the quantum mechanic says, on how—and when—you look at it. Common sense and aesthetics say it must be one or the other; the paradox bothers nonphysicists, and indeed it bothered Einstein. Yet the quantum mechanic is right. Elegant benchtop experiments show that the bizarre quantum world is real.

54 

The Very-Long-Baseline Array
Kenneth I. Kellermann and A. Richard Thompson

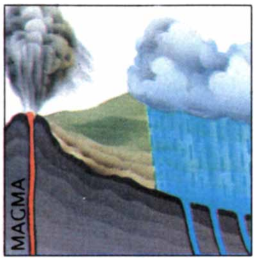
The resolution of a radio telescope depends on the diameter of its antenna. The VLBA, an array of 10 antennas stretching from the Virgin Islands to Hawaii, will have an effective "diameter" of some 8,000 kilometers. With it astronomers will study such deep questions as whether black holes are the source of the prodigious energy of quasars.

64 

Intertidal Fishes
Michael H. Horn and Robin N. Gibson

How can fish live in the harsh intertidal zone, where they are buffeted by waves and currents when the tide is in and are often left high and dry at low tide? Remarkable adaptations—including tolerance of dehydration, fins modified for crawling and clinging and even the ability to breathe air—make survival possible where the oceans and seas meet the land.

72



The Not-So-Rare Earths
Gunter K. Muecke and Peter Möller

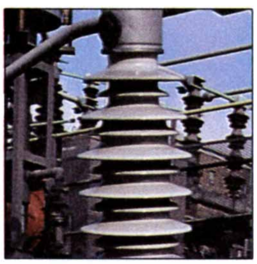
They are neither rare nor earths. They are metallic elements present in small concentrations in most minerals. Their relative abundance gives geologists clues to the geochemistry of magmas and hydrothermal solutions. The rare earths are essential constituents of advanced alloys, laser crystals and the new high-temperature superconductors.

78


Art, Illusion and the Visual System
Margaret S. Livingstone

The verve of op art, the serenity of a pointillist painting and the 3-D puzzlement of an Escher print derive from the interplay of the art with the anatomy of the visual system. Color, shape and movement are each processed separately by different structures in the eye and brain and then are combined to produce the experience we call perception.

86

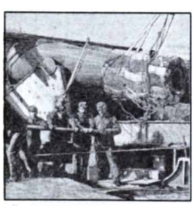

The Transformer
John W. Coltman

Invented in the mid-19th century, it is still essential: it makes electricity a convenient form of energy by converting a generator's low-voltage, high-current output into high-voltage, low-current power for efficient transmission—and then back again for consumption. Without it homes, office buildings and factories would need their own generators.

DEPARTMENTS

8 **Letters**

12


50 and 100 Years Ago

1888: An Army engineer has made an electromagnet from a pair of 20-ton cannons.

100


Computer Recreations

Atomic bearings, molecular computers and machines to ream out blood vessels.

14 **The Authors**

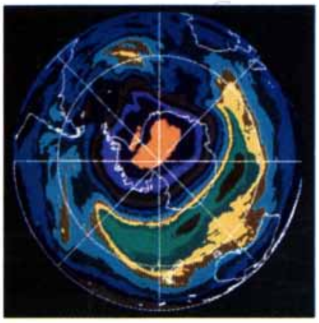
17 **Science and the Citizen**

96 **The Amateur Scientist**

104 **Books**

108 **Annual Index**

112 **Bibliography**



THE COVER shows a map, based on satellite data, of ozone levels in the atmosphere over the Southern Hemisphere on October 15, 1987 (see "The Antarctic Ozone Hole," by Richard S. Stolarski, page 30). The orange area is the heart of the ozone-depleted region, or ozone "hole," detected over Antarctica in the spring for the past several years. Recent studies support the hypothesis that chlorinated chemicals contribute to the hole, but other factors also seem implicated.

THE ILLUSTRATIONS

Cover image courtesy of National Aeronautics and Space Administration

Page	Source
31	NASA
32-34	George V. Kelvin, Science Graphics
35	Mark Muller, National Science Foundation
36	George V. Kelvin
38-39	Gilla Kaplan, Rockefeller University
40-41	John Ding-E Young, Rockefeller University
42-43	Dana Burns
44	Gilla Kaplan
47	Roger Gérard, Institut d'Optique Théorique et Appliquée
48-53	Gabor Kiss
55-60	George Retseck
61	R. C. Walker and J. M. Benson, National Radio Astronomy Observatory, and Stephen C. Unwin, California Institute of Technology
62	P. N. Wilkinson, Nuffield Radio Astronomy Laboratories, England (*top*); Robert L. Mutel, University of Iowa (*bottom*)
63	George Retseck
65	Tom Prentiss
66	Anne Wertheim
67	Anne Wertheim (*top*), Michael H. Horn (*bottom*)
68-70	Tom Prentiss
73-76	Ian Worpole
77	Gunter K. Muecke and Peter Möller (*top*), Ian Worpole (*middle and bottom*)
79	Margaret S. Livingstone
80	Patricia J. Wynne
81	Richard MacDonald (*top*), Margaret S. Livingstone (*bottom*)
82	David H. Hubel
83	The Bettmann Archive
84	Richard MacDonald
85	The Baltimore Museum of Art: The Cone Collection, formed by Dr. Claribel Cone and Miss Etta Cone of Baltimore, Md. (*top*); Patricia J. Wynne (*bottom*)
87	James Kilkelly
88-90	Hank Iken
92	Westinghouse Electric Corporation
93-94	Hank Iken
96-99	Michael Goodman
101	Carol Donner
102-103	Andrew Christie

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.

Before he pushes the edge, IBM takes you beyond it.

From miles away and thousands of feet below, you know exactly what he's feeling.

Using an IBM 3090 with supercomputer power, you've pushed your design through relentless iterations of finite element modeling and fluid dynamics analysis. You've defined the edge.

The IBM 3090 with Vector Facility, for its part, has redefined supercomputer performance, combining up to six vector processors with one of the industry's fastest scalar processors and most efficient memory management systems, parallel processing capabilities, and full IBM System/370 compatibility.

The result is balanced supercomputer performance that can improve your job turnaround by eliminating data bottlenecks, and save the added expense of a separate

front-end processor. Plus, the IBM Vector Facility's modular design lets you start with just the computer power you need today, and make economical in-field upgrades as your needs grow. And that gives the IBM 3090VF a big edge over more expensive and less flexible vector processors.

IBM's integrated system approach combines the power of the IBM 3090VF with the advantages of a single operating environment, low entry cost, low software and maintenance costs, a wide variety of application programs, and IBM support to create a winner in total cost-of-computing.

To find out more, or to arrange to have your IBM Marketing Representative contact you, simply call 1-800-IBM-2468, ext. 20. IBM®

LETTERS

To the Editors:

In "How Children Learn Words" [SCIENTIFIC AMERICAN, September, 1987] George A. Miller and Patricia M. Gildea present a provocative set of experimental findings, but I believe they miss one major point and arrive at a second in a roundabout way.

In the mid-1950's Noam Chomsky proposed that two kinds of grammar serve to organize English sentences: a generative grammar and a recognition grammar. The same concept can be applied to vocabulary. High school seniors who can recognize 80,000 words in a written or spoken context call on far fewer when they write their own sentences.

The reason is probably that (as Miller and Gildea point out) words are acquired in stages. Words do not ascend to the generative vocabulary until the "slow" second stage of acquisition. But in the work reported in the article children were asked to write sentences including new words soon after the words were introduced. These children will succeed only if the new words are close synonyms of ones they already know.

Instead children should be exposed to new words over a period of time, reading them in several contexts and being tested for recognition. Only when they have learned to distinguish nuances should they be asked to use the words in sentences.

The authors allude to the second major point at the end of the article. Behind each word lie several stories; the sum of the stories of all the words represents much of our common knowledge. The words themselves do not tell the stories; they only represent them. Thus the major task in learning words is to learn the stories behind them and fit the stories into the cognitive framework developing in the mind.

For example, a child cannot learn the meaning of *erode* without first learning a smattering of geology. Furthermore, many uses of words are metaphorical, and employing those meanings in sentences requires an understanding that goes well beyond the strict definition. The phrase "The president's popularity was eroded," for instance, uses geologic erosion as a metaphor for declining popularity.

And so while it is delightful to dream of a computer program that will regale a child with stories and patiently explain metaphors, I doubt that such a program will be written soon. Simply presenting timely but canned definitions on a computer screen will be no substitute for the reading and talking necessary to impart the rich knowledge that permeates every word.

DAVID POMERANTZ

Wakefield, Mass.

To the Editors:

Miller and Gildea leave us with an apparent contradiction. On the one hand they rightly emphasize the astonishing speed and apparent ease with which children, in the normal course of growing up, master a very large number of words. On the other hand they look for ways to help them do what they already do so well. If children learn words so easily, why do they need help?

Americans have an odd admiration for a big vocabulary. From popular advice on "How to Increase Your Word Power" or "Ten Minutes a Day to a Better Vocabulary" to the verbal-aptitude tests by which we sort students, a massive vocabulary is widely regarded as a Good Thing. But as Miller and Gildea sagely point out, words are best learned in their natural context, and this implies that they are acquired as a by-product of learning about subject matter. Even children who do poorly in school have no difficulty mastering the vocabulary of baseball or drugs or clothes if these happen to be the topics that grip them. Is it not our goal simply to get children excited about subject matter? Why worry about vocabulary when children learn words so easily as long as they are interested?

There is a long tradition in American education of trying to solve other problems by manipulating language. Hence we struggle to "improve" a child's pronunciation or grammar, and we are now offered a new way of "improving" vocabulary. The forms of language, whether they are vocabulary, grammar or pronunciation, are the result of an individual's experience, opportunities and social background, not their cause. Change the opportunities and language will adapt. Get children interested in more things and their vocabulary will grow. The hope that we can solve some other problems (whether they relate to social status or to intellectual interests) by focusing on language is largely illusory.

The question of how children learn so many words with apparent ease remains a fascinating one, and I find Miller and Gildea's suggestions about how they do and do not learn words both exciting and convincing. Nevertheless, the implication of their article is that some kind of problem will be solved if we can help children to learn words even faster. What is the problem? Is the goal merely to help them get a high mark on a verbal-aptitude test?

ROBBINS BURLING

Department of Anthropology
University of Michigan
Ann Arbor

To the Editors:

We are in general agreement with both writers. The letters indicate, however, that we did not distinguish clearly enough between the scientific problem of studying how children learn words and the practical problem of facilitating that learning.

Our basic interest, and the central message of our article, is scientific, even though our results do have obvious implications for the practical problems of language instruction. For example, our research establishes that children often misunderstand dictionary definitions, and so we certainly agree with David Pomerantz that an educational policy of "simply presenting timely but canned definitions on a computer screen" would not be an effective way to teach. Again, our research shows that children who are asked to write sentences containing unfamiliar words often fail. We have argued that the sentence-writing task is a valuable research tool because the way children fail tells us interesting things about how they learn. But we certainly do not advocate the task as a method for teaching vocabulary; we agree with Pomerantz that other approaches are preferable.

When Robbins Burling asks why we look for ways to help children do what they already do so well, his criticism is not aimed at our research or interpretations. His concern seems to be that the introduction of computers adulterates a natural learning process that is better left alone. We agree with Burling that the natural process can be highly effective, but we doubt that it is always given a chance.

Our country spends vast sums on education so that all children will learn to read and write, because literacy is essential to becoming a productive member of our complex information economy. We agree with Burling that literacy and vocabulary growth go together naturally—reading comprehension depends on a large vocabulary, and large vocabularies result from extensive reading. The natural learning process breaks down, however, when children read little or nothing outside school. In school they probably read fewer than one million words a year, which is simply not enough for much more than a conversational vocabulary. It is difficult to change the social and economic conditions under which many children live in noisy, crowded places with no privacy for study, own no books of their own and have parents who are illiterate or often absent. But we might be able to do something with computers that would make the reading they do in school more effective.

These comments, of course, go well beyond the scientific observations our article was meant to report. Perhaps it was our relative neglect of such practical topics that led the writers to question the educational relevance of our approach.

GEORGE A. MILLER

PATRICIA M. GILDEA

ERRATA

In "Interfaces for Advanced Computing," by James D. Foley (October, 1987), the force-input system developed at George Washington University should have been attributed to Teresa W. Bleser. Owing to an editorial error the last paragraph of the article implies that experiments done by the author and his colleagues were the first to assess the value of artificial realities since the work of James J. Batter in the 1960's; actually there had been many related studies in the interim.

In "The Ancestry of the Giant Panda," by Stephen J. O'Brien (November, 1987), an editorial change introduced an error on page 106. The date of 35 million years ago is given for the divergence between human beings and African apes; actually the date applies to the divergence between humans and Old World *monkeys.*

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JANUARY, 1938: "There is much discussion about whether the passage over the junction point between a nerve and a muscle cell is an electrical or a chemical process. There is much evidence to show that the transition is effected by chemical transmitters, such as acetyl choline, in the case of our voluntary and involuntary movements."

"The captive balloon, long familiar as a wartime device for keeping watch over enemy activities from the air, has been given an important peacetime job—that of surveying city traffic. In this new use it will enable traffic engineers to study at their leisure actual conditions at key locations and draw conclusions as to safety measures that should be put into practice or changes in traffic regulations that should be made. As a traffic surveyor, the captive balloon is only 12 feet in diameter and carries a camera, which is operated by remote control from the ground."

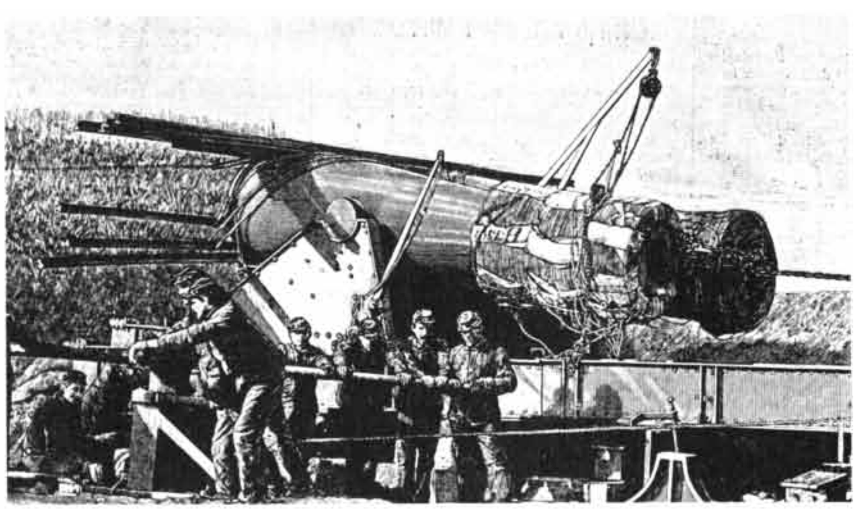
"The stonecutter of old, hacking at hard rock with a pick or with hammer and chisel, had little to fear from dust. The modern stonecutter, thundering through brittle rock with power drills, stirs up smokelike waves of dust. Unless properly protected, he may inhale harmful quantities of dangerous silica particles. Some of these drills are known among miners as 'widow makers.'"

"Increased use of tar on roads throughout the country is causing the water supplies of hundreds of cities to take on objectionable tastes and odors, reports the American Institute of Sanitation. Road tar contains small amounts of phenolic chemicals, which are leached out by the rain and carried along to the lakes, rivers and reservoirs from which cities obtain their water supplies. The chemicals are usually present in very small amounts and ordinarily are unnoticeable to the taste. But when the water is chlorinated, the phenolic substances are turned into pungent compounds having a pronounced medicinal taste."

"Homes of the future may be illuminated by a network of transparent quartz rods that transmit the light from a single bulb, suggests a scientist. The quartz rods would act as pipes, the light 'flowing' through them as does water through an iron pipe."

SCIENTIFIC AMERICAN

JANUARY, 1888: "The success of the anti-vaccinationists has been aptly shown by the results in Zurich, Switzerland, where for a number of years, until 1883, a compulsory vaccination law obtained, and smallpox was wholly prevented—not a single case occurred in 1882. This result



Major King's great cannon magnet

was seized upon the following year by the anti-vaccinationists and used against the necessity for any such law, and it seems they had sufficient influence to cause its repeal. The death returns for that year (1883) showed that for every 1,000 deaths two were caused by smallpox; in 1884 there were three; in 1885, 17, and in the first quarter of 1886, 85."

"A locomotive possessing several unusual features has been recently built by the Hinkley Locomotive Company, of Boston. The most novel feature is the form of the tread of the driving wheels. The circumference of the tire, instead of being a true circle, is polygonal and formed of 105 flats, each about two inches long. The object is to prevent slipping."

"The mastodon, that great fossil mammal, allied somewhat nearly to the elephant, has become perhaps more familiar to the public than any other of the numerous great creatures that once lived in our extended country. In nearly every state west of New England portions of this creature have been disinterred. The circumstance of several examples' having about them evidences of man's work is extremely interesting. We are therefore able to say that man and mastodon are contemporaneous. But the date is obscure. We have not determined what sort of man made the stone arrowheads that struck the life out from the great carcasses and lie among their remains."

"Dr. Victor C. Vaughan, Professor of Physiological and Pathological Chemistry at the University of Michigan, has announced the result of a series of experiments, which have ended, he claims, in the confirmation of the germ theory in cases of typhoid fever. The fever was produced in a cat with more completeness and success than has ever before attended such endeavors."

"Mr. W. P. Gebhard tells how those who are neither plumbers nor sanitary inspectors may locate the slightest leakage in water pipes by introducing essence of peppermint into them."

"An interesting magnetic experiment on a large scale has been made at Willets Point, N.Y., by Major W. R. King, of the Engineer Corps, U.S.A., consisting in the conversion of a pair of great cannons, each weighing over 20 tons, into an electromagnet."

SCIENTIFIC AMERICAN

CORRESPONDENCE

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

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.25 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/88. $1.25 + 0.00.

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

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

Address subscription correspondence to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 953, Farmingdale, N.Y. 11737. Telephone inquiries: 1-800-227-5782. The date of the last issue on subscriptions appears at the right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name ______

New Address

Street ______

City ______

State and ZIP ______

Old Address

Street ______

City ______

State and ZIP ______

THE AUTHORS

RICHARD S. STOLARSKI ("The Antarctic Ozone Hole") is a research scientist in the atmospheric chemistry and dynamics branch of the National Aeronautics and Space Administration's Goddard Space Flight Center. After earning a doctorate in physics from the University of Florida in 1966, he accepted a research position at the University of Michigan. He left there in 1974 to join NASA as a research physicist at the Johnson Space Center. Stolarski moved to Goddard in 1976, where he was head of the atmospheric chemistry and dynamics branch from 1979 to 1985.

JOHN DING-E YOUNG and ZANVIL A. COHN ("How Killer Cells Kill") work together in Cohn's laboratory of cellular physiology and immunology at Rockefeller University. Young, who is assistant professor at Rockefeller, was born in Taiwan but is a Brazilian citizen and a permanent U.S. resident. He obtained an M.D. in 1979 at the National University of Brasilia and then completed a doctorate and a two-year fellowship at Rockefeller before assuming his current position in 1985. Cohn is professor and senior physician at Rockefeller and adjunct professor of medicine at the Cornell University Medical College. He got his M.D. at the Harvard Medical School in 1953 and did his internship and residency at the Massachusetts General Hospital. After serving as chief of the division of rickettsial biology at the Walter Reed Army Institute, he went to Rockefeller in 1958. Cohn was codirector of the joint Rockefeller University-Cornell University Medical College M.D.-Ph.D. program from 1973 to 1978.

ABNER SHIMONY ("The Reality of the Quantum World") is professor of philosophy and physics at Boston University. He received a B.A. (1947) and a Ph.D. in philosophy (1953) from Yale University and a Ph.D. in physics (1962) from Princeton University. He taught at the Massachusetts Institute of Technology until 1968 and has held visiting positions at the University of Paris (Orsay), Mount Holyoke College and the Swiss Federal Institute of Technology in Zurich. Shimony specializes in the foundations of quantum mechanics and is particularly interested in the design of experiments that test philosophically motivated questions. He is active in the peace movement and writes poems and light plays, some of them relevant to physics.

KENNETH I. KELLERMANN and A. RICHARD THOMPSON ("The Very-Long-Baseline Array") are colleagues at the National Radio Astronomy Observatory in Charlottesville, Va., which is operated by Associated Universities, Inc., under contract with the National Science Foundation. Kellermann, a senior scientist at the observatory, got a bachelor's degree at the Massachusetts Institute of Technology in 1959 and a Ph.D. in physics and astronomy from the California Institute of Technology in 1963. He has held visiting positions at Caltech and also in Australia and Germany. Thompson is deputy project manager and head of electronics for the Very-Long-Baseline Array (VLBA). He received his undergraduate (1952) and doctoral (1956) degrees from the University of Manchester, where he developed a radio interferometer in 1956. After coming to the U.S. in 1957 he did radio astronomy at Harvard and Stanford universities. Thompson joined the observatory in 1973 and worked on the design and construction of the Very Large Array before turning to the VLBA.

MICHAEL H. HORN and ROBIN N. GIBSON ("Intertidal Fishes") have done joint fieldwork in the Mediterranean and on the Atlantic coast of France. Horn is professor of biology at California State University at Fullerton. He earned a master's degree at the University of Oklahoma in 1965 and a Ph.D. from Harvard University in 1969 and went to Fullerton after spending a postdoctoral year at the Woods Hole Oceanographic Institution and the British Museum (Natural History). Gibson, principal scientific officer at the Scottish Marine Biological Association in Oban, got his Ph.D. in 1965 from the University of Wales and became interested in intertidal fishes as a postgraduate student there. He has continued his studies of the fishes both in Europe and in the Middle East, with occasional forays to the Pacific. Gibson was recently awarded a D.Sc. for contributions to the biology of shallow-water fishes.

GUNTER K. MUECKE and PETER MÖLLER ("The Not-So-Rare Earths") have collaborated since 1983, studying the distribution and mobility of trace elements, including the rare earths. Muecke is associate professor in the departments of geology and environmental studies at Dalhousie University in Halifax. After receiving bachelor's and master's degrees in geology at the University of Alberta in 1964, he obtained a doctorate from Oxford University in 1969 and then went to Dalhousie the following year. Muecke has also been a visiting investigator at the University of Göttingen and the Hahn-Meitner Institute of Nuclear Studies in West Germany. Möller is senior scientist at the Hahn-Meitner Institute. He was awarded a doctorate in nuclear chemistry in 1967 by the Technical University of Berlin and began teaching soon after, both there and at the Free University of Berlin. He continues to teach at both institutions.

MARGARET S. LIVINGSTONE ("Art, Illusion and the Visual System") is associate professor of neurobiology at the Harvard Medical School. After receiving her undergraduate degree at the Massachusetts Institute of Technology, she entered the Harvard Medical School, but she transferred to the graduate school and earned a Ph.D. there in neurobiology in 1980. Since 1979 she has collaborated with David H. Hubel of the medical school in the study of how human beings and other primates process visual information. Livingstone has also studied the roles of neurohormones in the behavior of lobsters and the biochemical basis of associative learning in *Drosophila*.

JOHN W. COLTMAN ("The Transformer") is the retired director of research and development at the research laboratories of the Westinghouse Electric Corporation. He is best known for the invention and development of the X-ray image amplifier, a device that is now standard in hospital radiology departments. Coltman earned his bachelor's degree in physics at the Case Institute of Technology in 1937 and his doctorate in nuclear physics from the University of Illinois in 1941. He then joined Westinghouse and did research on microwave tubes for radar and for jamming enemy radar. In 1949 he was named manager of the electronics and nuclear physics department. He held two other positions before accepting his last title in 1974. Coltman, who left Westinghouse in 1980, is a flutist and has published many articles on acoustics.

SCIENCE AND THE CITIZEN

18 PHYSICAL SCIENCES
21 BIOLOGICAL SCIENCES
23 MEDICINE
24 TECHNOLOGY

Twice Burned

Controversy erupts—again—over a key "Star Wars" weapon

Although the X-ray laser has yet to show that it can shoot down a missile, it has demonstrated a capacity for burning proponents of the Strategic Defense Initiative. President Reagan, for example, has declared that the initiative would make nuclear weapons "impotent and obsolete," yet the X-ray laser is itself a nuclear weapon. Unclassified reports describe it as a nuclear bomb packed in an array of metal rods; as the bomb explodes, the rods channel its X-ray emission into coherent beams intense enough to destroy missiles thousands of miles away.

The weapon has a powerful patron in Edward Teller, renowned as the father of the hydrogen bomb and a founder of the Lawrence Livermore National Laboratory. Teller is credited with convincing the president that an SDI system in general and the X-ray laser in particular are feasible. Research into the X-ray laser and related "nuclear-pumped" weapons consumes a substantial slice (up to $600 million) of the SDI pie year after year. Much of the money is spent exploding experimental devices underground at the Nevada Test Site.

Disputes over the technical status of the X-ray laser first became public three years ago. Sources at Lawrence Livermore, which leads the weapons laboratories in developing the technology, told the press that an underground test in March, 1985, had produced X rays more than a million times brighter than those produced in earlier tests. These anonymous leaks (the technical status of the program is highly classified) provoked anonymous rebuttals from sources who said the instrumentation used for measuring the X rays was flawed. Members of Congress called for a halt in tests of the device, but testing continued after an investigation of the dispute by the General Accounting Office, an arm of Congress, proved inconclusive.

The X-ray laser has recently stirred up more trouble. Two physicists who have worked on the program—Roy D. Woodruff, a former associate director, and W. Lowell Morgan, a staff scientist—have charged that Teller and other officials at Lawrence Livermore misled officials in Washington about the status of the research. Woodruff was an unwilling whistle-blower. Last spring he wrote a confidential letter to the University of California, which manages Lawrence Livermore for the Department of Energy, contending that Teller and Lowell L. Wood, Jr., manager of a key group of weapons designers at Lawrence Livermore, conveyed "overly optimistic, technically incorrect statements regarding this research to the Nation's highest policy makers."

Woodruff maintained further that the laboratory's director, Roger E. Batzel, "refused to transmit correcting technical information or allow me to do so." After he asked to be reassigned, Woodruff said, Batzel demoted him and gave him minimal salary increases. Woodruff's complaints were made public after the Southern California Federation of Scientists obtained a copy of his letter from an anonymous source and released it to the press.

Spokespersons for Lawrence Livermore and the University of California have dismissed the dispute as a "personnel matter" and a "legitimate difference of opinion between reputable scientists." Teller, Wood and Batzel have declined further comment. But Woodruff's account of official misrepresentation was recently corroborated by Morgan, who worked in the X-ray laser program from 1981 through 1985, when he asked to be reassigned. Morgan, now on leave from Lawrence Livermore and a fellow at the University of Colorado at Boulder, aired his views in a letter to Representative George E. Brown, Jr., of California. Tests of the X-ray laser, Morgan contended, have yielded "miniscule returns of poor quality data (by conventional standards in physical science) for a tremendous investment of effort and, in my opinion, an obscene investment of money." The "few scientific results that we had," he said, "were being grossly misrepresented." Morgan also asserted that experts from the General Accounting Office and from JASON, a group of consultants to the Defense Department, when briefed on the program, saw only data that had been "massaged and 'made pretty.'"

In separate interviews with SCIENTIFIC AMERICAN, Woodruff and Morgan both suggested that Reagan's announcement of the SDI in 1983 had placed undue pressure on what had been only moderately promising research. "After the president's speech," Woodruff said, "Teller and Wood's representations assumed an almost frenetic character." Morgan noted that "in the hype that followed Reagan's speech it was clear that it was becoming a boondoggle."

Woodruff has been at least partially vindicated: the University of California recently ruled that he was unfairly demoted and ordered the laboratory to give him a better position. The General Accounting Office is also conducting an investigation of his charges. —*John Horgan*

The Crash

Will Wall Street's woes hurt higher education?

Crash. Crack. Correction. Whatever it was that befell the stock market in October, it has left officials in higher education anxious about the future of their institutions. Summing up the concern, Anthony Cerami, dean of graduate students at Rockefeller University, remarks, "Everyone cuts back on research and education in hard times."

The impact on many schools' endowments was immediate and dramatic. On October 19, the "Black Monday" when the Dow Jones average plunged more than 500 points, the value of investments in Stanford University's endowment fell by $200 million, more than 10 percent. The $3.8 billion Texas Permanent University Fund, which supports the University of Texas and Texas A. & M. University systems, lost $400 million in value. Some university officials, pointing out that by the end of November the market stood about where it had a year earlier, dismiss these declines as "paper losses" that will have little effect on their long-term finances. "Our endowment has declined, but it's higher than it was at the start of the year," says Glenn P. Strehle, treasurer of the Massachusetts Institute of Technology. "The increase didn't have much effect, and the decrease didn't either."

Others are less sanguine. Rockefeller University, whose $560 million endowment fell by $45 million, recently trimmed the percentage of its endowment allocated to the annual budget. "It used to be 6 percent," says Barry W. Dress, Rockefeller's treasurer. "Now it will ratchet back to 5 percent. There's just a little bit less for new initiatives and equipment. It's marginal, but it's corrosive."

Roderick MacDougall, the treasurer of Harvard University, thinks "the biggest area of concern" is that people who have lost money in the stock market, or who fear future losses, will curb their gift giving. Donations from wealthy individuals, he says, "constitute a major portion of what makes American higher education superior." Fund raising had already been dampened by legislation enacted in 1986 that reduces the tax deductions allowed for gifts. "People are having second thoughts about their gifts now," observes Theodore P. Hurwitz, vice-president for institute relations at the California Institute of Technology. "They will defer them or put them in their wills."

David F. Mears, coordinator of contracts and grants for the University of California system, thinks private foundations, many of which have invested heavily in stocks, may also curtail their donations. "I have a friend who is the president of a foundation that just lost $40 million," he says. "That will make him look a lot harder at spending."

Large corporate benefactors such as IBM and Monsanto—even if their earnings are not directly affected by a drop in the value of their own stock—may reduce research grants to universities out of fear that hard times are coming, according to Dress. "They will certainly cut grants to universities before they cut their own internal R&D," he says. Many economists have predicted that Black Monday foreshadows a recession, during which consumer and business spending will shrink markedly. James Rosse, provost of Stanford and an economist, notes that no one can predict a recession with certainty. "But I think it is fair to say," he adds, "that the probability that we will slip into a recession over the next 12 months has increased."

The Federal budget deficit has been widely blamed for the stock-market slump; the budget-cutting efforts spurred by the slump may mean a decline in grants to university investigators from Government agencies. Mears notes that the Gramm-Rudman-Hollings Act, which stipulates a reduction in the 1988 budget of $23 billion, would have decreased the funds available for research anyway. "If both political parties feel that more cuts are needed to bolster confidence in the economy," Mears remarks, "the cuts may be larger and they may be projected further into the future." —*J.H.*

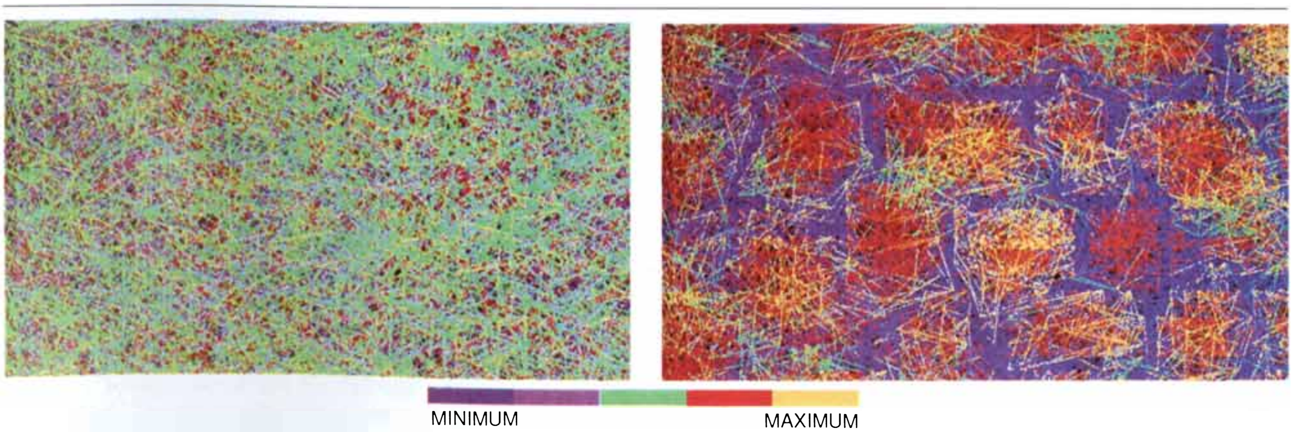
"Neural Darwinism"

Competition among neurons is simulated on computers

The brain contains maps that represent sensory information, so that when a specific receptor, such as a cell in the retina, is stimulated, it triggers a response in a specific and highly localized area of brain cells. The orderly mapping follows a genetic blueprint, but its details are fine-tuned by experiences early in life and can evolve throughout adulthood. What kind of wiring scheme could account for such maps, at once structured and yet flexible?

One theory, proposed by Gerald M. Edelman of Rockefeller University, suggests that the brain has the potential to represent vast numbers of things, but selects only those that are actually experienced. Such selection can occur, says Edelman, if the basic unit of cortical organization is not the single neuron but a "neuronal group." The cells in a group "vote" among themselves to represent a particular area of sensory receptor cells, called a receptive field. Unfortunately neuronal groups, if they exist, would be only about 50 micrometers wide, too small to be studied by current methods. And so Edelman, Leif H. Finkel and John C. Pearson have used computer simulations to show how a network with realistic anatomical and biochemical features

Clusters of neurons may form the basic unit of sensory data in the brain



DENSE WEB OF CONNECTIONS with random strengths links 1,536 neurons in a simulation of a cortex (left). *Stimuli from a "hand" cause the network to form clusters of strongly connected cells surrounded by pale blue moats of weak connections* (right).

can spontaneously give rise to neuronal groups. The work is described in *The Journal of Neuroscience* and in Edelman's book, *Neural Darwinism*.

Edelman and his colleagues have simulated part of the somatosensory cortex, a region that receives signals from sensory nerves in the skin in such a way that anatomical relations are preserved. For example, a patch of neurons that respond when the index finger is touched lies between patches of neurons that respond to the thumb and middle finger. In the simulation the workers mathematically modeled a sheet of cortical cells connected to the surface of the hand. When a receptor in the hand receives a stimulus, it transmits excitatory signals along "nerves," which fan out to a patch of cortical cells that are densely interconnected.

The simulation modifies the efficiency with which signals get passed between two neurons—their "synaptic strength"—based on the activity in the network. Repeated experiences will strengthen some synapses and fortify certain patterns of activity. The synaptic-modification rule used in the simulation dictates that the strength of a synapse increases if there are both an input at the synapse and a large number of additional inputs being conducted from the other synapses of the cell.

At the start of e ch simulation the synaptic stre ıgths are set at random. When the hand is "touched," a cluster of receptors sends signals to a patch of cortical cells. There the cells that receive the inputs and are also heavily interlinked act on one another through positive-feedback loops to strengthen their connections, resulting in clusters of strongly connected cells. Cells that are not stimulated at the same time or that lack sufficient connections to the group do not join the cluster. The same process that led to the formation of the clusters also leads each cluster to develop strong connections to a particular receptive field.

Remarkably, the simulated cortical representations that result closely resemble those observed in animals, particularly by Michael Merzenich of the University of California at San Francisco. The simulation, for example, produced groups that chose a receptive field exclusively on either the palm or the back of the hand, although all cells receive equal numbers of nerve connections from both sides. When one finger was tapped repeatedly, the groups that represent it grew in size and number. When certain "nerves" were severed, some groups became inactive and others switched over to represent other areas of the hand. *—June Kinoshita*

PHYSICAL SCIENCES

Phasing In

New X-ray-diffraction method reveals molecules' 3-D structure

X-ray crystallography, which enables workers to deduce the atomic structure of a crystalline material by observing how it diffracts X rays, is indispensable to the study of complex organic molecules. It is also extremely difficult. Recently, for example, a team of six workers from Harvard and Stanford universities finally unraveled the convoluted structure of the histocompatability antigen HLA-A2, a molecule crucial to the immune system, after years of probing crystalline specimens.

The major problem lies in extracting three-dimensional information from the spots produced when diffracted beams strike a film or other detecting medium. The intensity of each spot and its position on the film reveal much about the identity and position of the atoms that diffracted the beam. But to model the crystal precisely, workers also need to know the phase (crest, trough or intermediate point) of the diffracted beam's waveform as it strikes the film.

A technique developed by Roberto Colella of Purdue University and Qun Shen of Cornell University may provide investigators with a powerful new tool for determining phase. The two workers exploit a phenomenon that is well known to crystallographers: if a crystal is oriented in a certain way, the beam simultaneously diffracts off not one plane within the crystal but two and splits into two overlapping beams.

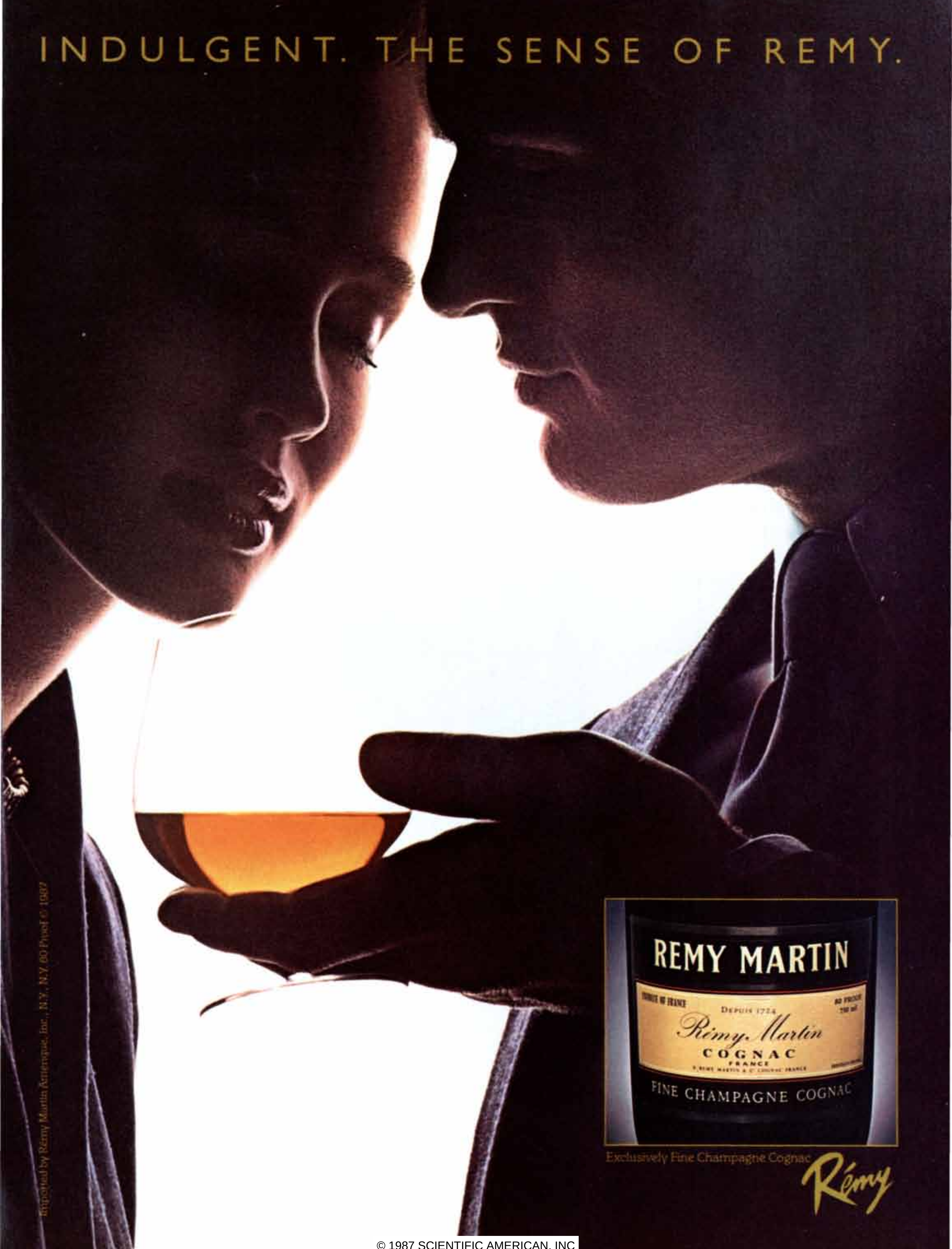
To carry out their procedure, Colella and Shen orient the crystal so that it produces a single diffracted beam. Then they rotate the crystal to the position in which the beam splits. If the two beams are in phase—crest aligned with crest and trough with trough—they reinforce each other and the spot they produce on the film brightens; if the beams are out of phase, they cancel each other and the spot dims or disappears. By measuring the change the workers can calculate the phases of the two overlapping beams with respect to each other; by repeating the process with the crystal oriented in many directions, the workers eventually gather enough data to unravel its three-dimensional structure.

Although investigators first considered exploiting multiple-beam diffraction decades ago, only in recent years have advances in synchrotron X-ray sources and computers for analyzing data made the procedure feasible. More advances are needed: so far the method has successfully analyzed only relatively small organic molecules, such as benzine. Further refinements, other investigators say, could make the procedure capable of probing molecules that resist analysis by the two techniques now most commonly used.

One of the techniques, called isomorphous replacement, involves inserting a heavy atom, such as gold or lead, into the molecule. X rays strongly diffracted by the heavy atom interfere with—that is, alter the phase of—the X rays diffracted by the lighter organic atoms (generally hydrogen, oxygen and carbon). By "decorating" the same molecule with different heavy atoms and then comparing the patterns of interference, workers can deduce the position of the light atoms in the molecule.

So-called direct methods, developed in the late 1950's by Jerome Karle of the U.S. Naval Research Laboratory and Herbert A. Hauptman of the Medical Foundation of Buffalo, use statistical analysis to determine phase. The analysis exploits relations among the spots produced on a film as a crystal rotates through different positions; for example, if the spatial coordinates of three spots with similar intensities add up to zero, their phases should cancel one another. Karle, who shared a 1985 Nobel prize with Hauptman, notes that since the direct methods incorporate probability theory, they occasionally produce erroneous results. If direct-methods analysis can be supplemented with "definitive" data from multiple-beam tests, Karle says, "that would be very helpful."

Karle points out, moreover, that the direct methods cannot reliably reveal the structure of molecules consisting of more than about 100 atoms. Isomorphous replacement, on the other hand, works well only for


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

molecules that have more than about 400 constituent atoms; smaller molecules absorb a heavy atom with difficulty, often at the price of structural distortion. If Colella and Shen can apply their experimental technique to molecules in the intermediate "window," Karle observes, "that would be a great leg up." —*J.H.*

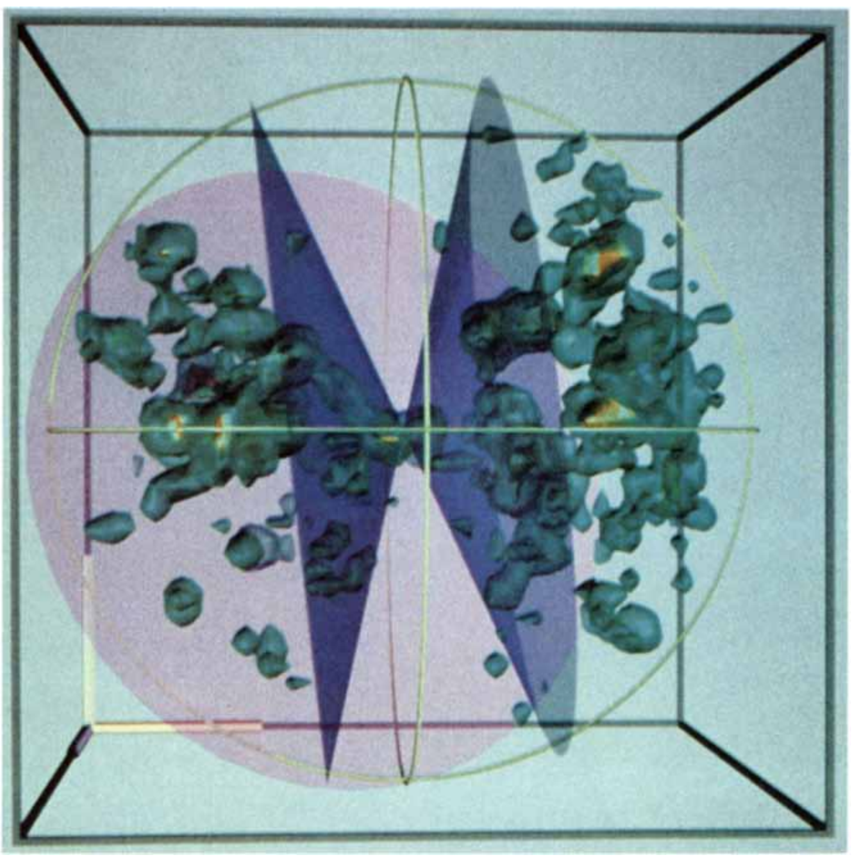
Cosmic Complex

Map of galaxy clusters reveals the largest structure yet seen

The dimensions of our cosmic address continue to grow. Astronomy began by mapping the earth's immediate neighborhood: the solar system. Early in this century workers established that the solar system inhabits a spiral-shaped island of stars, the Milky Way. It in turn was found somewhat later to be one of about 20 relatively nearby galaxies known as the Local Group. In the past decade investigators concluded that the Local Group resides near the edge of the Local Supercluster, a lumpy sheet of galaxies some 100 million light-years across that has as neighbors the Hydra-Centaurus Supercluster and perhaps a vast concentration of galaxies known as the Great Attractor (see "The Large-Scale Streaming of Galaxies," by Alan Dressler; SCIENTIFIC AMERICAN, September, 1987).

Now R. Brent Tully of the University of Hawaii at Manoa reports in *The Astrophysical Journal* that the Local Supercluster and its neighbors all belong to a much larger "complex" of superclusters. The structure is by far the largest ever delineated: it is at least a billion light-years across, about a tenth the span of the observable universe, and 200 million light-years thick. Its estimated mass is one quintillion (10^{18}) times that of the sun. Tully calls the structure the Pisces-Cetus Supercluster Complex because, observed from the earth, it appears to be concentrated in the constellations Pisces and Cetus.

The universe appears lumpy on a scale too vast for current cosmological theories to explain



"COMPLEXES" OF GALAXIES *populate a sphere two billion light-years across in a map made with a supercomputer by R. Brent Tully. The Pisces-Cetus Supercluster Complex is the horizontal structure; three other supercluster complexes—Aquarius, Hercules-Corona Borealis and Leo—appear respectively at the upper left, upper right and lower left. The empty wedges represent regions that have been obscured by dust in the Milky Way.*

While noting that more observations are needed to corroborate the report, Vera C. Rubin of the Carnegie Institution of Washington calls it "quite exciting." "It's very difficult work, and it takes a lot of nerve," she says. Tully mapped not individual galaxies but clusters containing about 1,000 galaxies each that had been observed and catalogued by George Abell in the 1950's. Tully estimated the distances of the clusters by gathering recent measurements of their red shifts. (The degree to which light from a galaxy is shifted toward the red end of the spectrum is assumed to be roughly proportional to its distance from the earth.)

Tully confined his mapping to a cosmic sphere with a radius of one billion light-years centered on the Milky Way; observations of galaxies outside the sphere are "horribly incomplete," he explains. After plotting the positions of Abell clusters inside the sphere, Tully found that most of them seem to be clumped into five distinct structures occupying only .5 percent of the sphere's volume. Filling the center of the sphere is the Pisces-Cetus Supercluster Complex. Two other supercluster complexes of comparable size, which Tully calls Hercules-Corona Borealis and Aquarius, float nearby. The two remaining complexes, Leo and Ursa Major, are smaller and hover near the sphere's perimeter; Tully describes them as "just poking their noses into the room."

Even before Tully's report, observations of dense collections of galaxies surrounded by huge regions devoid of visible matter were proving difficult for cosmologists to explain. "We have always thought that if you looked out far enough," James P. Wright of the National Science Foundation says, "things would smooth out. But that hasn't proved true."

A major piece of evidence for the big-bang account of the universe's origin is the faint microwave radiation that reaches the earth at almost precisely equal intensities from every direction; it is thought to be the faint afterglow of the big bang itself. Theorists maintain that the seeds of large-scale structures such as the Local Supercluster and now Tully's complexes must have been present at the earliest stages of the universe's expansion; the microwave background radiation, they say, should show corresponding irregularities.

Probably the major attempt to explain the discrepancy between the smoothness of the microwave radiation and the clumpiness of galaxies suggests that visible objects represent only a fraction of all the matter in the universe; when so-called dark matter is included, proponents say, the mass of the universe is smoothly distributed. "Some people describe the light-emitting matter as being just froth on top of a dark beer," Wright notes. There are many candidates for the role of dark matter—black holes, neutrinos and cosmic strings are the leading contenders—but no firm observational evidence supports any of them. Asked how he thinks his work can best be reconciled with the smooth microwave background, Tully responds, "I don't worry about that. I just worry about the reality of these structures." —*J.H.*

BIOLOGICAL SCIENCES

Optimalists under Fire

A philosophical dispute raises hackles among biologists

As they graze or hunt, many animals behave as though they were trying to maximize their total energy gain. For example, the great tit, a common European songbird, tends to abandon a waning food source when its rate of energy intake drops below what it could obtain by moving on. If choice prey are common, the bluegill sunfish avoids second-rate foods that would lower its overall rate of energy intake.

Such behavior conforms with some of the predictions of Optimal Foraging Theory (OFT), a prominent example of mathematical optimality arguments applied to biology. Yet OFT has been fiercely attacked. A recent salvo from Graham J. Pierce and John G. Ollason of the University of Aberdeen was published in *Oikos*, a Scandinavian journal of ecology. They titled their paper "Eight Reasons Why Optimal Foraging Theory Is a Complete Waste of Time."

OFT is founded in orthodox Darwinism. It assumes that eons of natural selection have produced animals whose behavior tends to maximize their fitness, loosely defined as each individual's ability to pass on its genes to subsequent generations. Fitness is hard to measure, and so workers instead study behavior in terms of some more convenient measure; in OFT the measure is usually the average rate of energy gain per unit of foraging time. The rationale is that better foragers are likely to leave more offspring.

Mathematical models taking into account such factors as the caloric content of food items and their distribution serve for calculating the best foraging strategy, as gauged by the chosen measure. If animals' actual behavior conforms to the prediction, the utility of the measure and the assumptions in the model are regarded as supported; otherwise the model may be revised. Finding that optimal-foraging models can predict behavior in many species, advocates believe, would testify to the broad explanatory power of natural selection.

Most of the objections Pierce and Ollason voice were originally put forward by Stephen Jay Gould and Richard C. Lewontin of Harvard University. These critics assert that foraging behavior is shaped by many factors in addition to the need to find food—watching for predators, for instance. Furthermore, behavior may still be evolving: environments and animals change over millenniums. Indeed, an "optimal" trait may never emerge, because genes originate by random mutation and interfere with one another in complex ways. All these factors mean that even if a model predicts behavior well, its success might be just coincidence.

John R. Krebs of the University of Oxford, one of the foremost OFT advocates, argues that the optimizing effects of natural selection are pervasive: "You only have to look around at things in nature, even bits of your own body, to see how incredibly beautifully designed they are to do the job they do." He and David W. Stephens of the University of Massachusetts at Amherst have written in their book *Foraging Theory* that the critics' objections "are reasons why optimality theorists might be wrong, but not why they are bound to be

Can models of "optimal" foraging explain how animals from sunfish to songbirds really act?

BOX TURTLE snacks on blackberries in a photograph from Animals Animals. Theories that describe foraging strategies in terms of orthodox Darwinism are under attack.

wrong." Krebs does think that some early OFT work was overambitious: "In hindsight it's not a law of nature but a way to explore what animals are doing."

Ollason says that animals are "obviously doing sensible things," and he does not dispute the power of natural selection. He nonetheless thinks claiming to know what behavior is optimal is "self-delusion" because of the enormous number of factors involved. He and Pierce argue that because the predictions of optimality models are seldom met exactly, the models are ipso facto wrong. Stephens replies that an impressive majority of studies of food choice and foraging strategy find at least qualitative agreement with the predictions, and that clear disagreements are few.

Lewontin acknowledges that optimality explanations in biology may sometimes be correct, but he complains that simple studies appearing to support OFT have been accepted uncritically when all they do is "validate one's ingenuity" at devising plausible models. Because demonstrating optimality convincingly is very difficult, he wonders whether evolutionary biology will have to limit itself to revealing the variety in nature without explaining particular events. Philip Kitcher, a philosopher at the University of California at San Diego, is more optimistic: the fate of OFT will depend, he says, on whether simple models work for a range of species or must be endlessly tinkered with. He suggests that "we are not yet in a position to make that distinction." *—Tim Beardsley*

The Nautilus and the Bomb

Fallout from nuclear tests reveals a nautilus's history

In *The Chambered Nautilus* Oliver Wendell Holmes wrote: "Year after year beheld the silent toil / That spread his lustrous coil; / Still, as the spiral grew, / He left the past year's dwelling for the new."

Holmes was guessing. Until recently no one knew exactly how long it takes a nautilus to grow a new chamber or to mature. The animal resists observation: its range is limited to tropical waters in the western Pacific and it usually swims from several hundred to more than 1,500 feet deep, below the range of human divers; when captured, it rarely survives for a year. Investigators have now unraveled the history of the nautilus, or at least one nautilus, contaminated by carbon 14 from nuclear explosions. By measuring the unnaturally high levels of the isotope in its shell, the workers found that the nautilus took about 12 years to mature; as it neared maturity it grew new chambers more slowly, but never at the rate of only one chamber a year, as Holmes had surmised.

Neil H. Landman of the American Museum of Natural History, Ellen R. M. Druffel of the Woods Hole Oceanographic Institution, J. Kirk Cochran of the State University of New York at Stony Brook and A. J. T. Jull of the University of Arizona began their inquiry by establishing the extent to which nuclear explosions in the 1950's and 1960's contaminated the environment of the nautilus. After the U.S. and the U.S.S.R. stopped testing in the atmosphere in 1963, fallout from the explosions continued to settle into the oceans. Banded coral from Australia's Great Barrier Reef, which grows a new layer every year, provided the investigators with a gauge of fallout contamination in the southwest Pacific. Beginning in 1958 the level of carbon 14 in successive bands was found to climb sharply above the natural background levels, peaking in the mid-1970's.

After it had established this background chronology, the team measured the carbon 14 in different chambers of a nautilus that was netted near New Caledonia, 750 miles east of Australia, in 1969 and died soon after. The shell has 31 chambers in all, and its markings indicate that the animal had reached sexual maturity, at which time the nautilus stops adding chambers. By means of accelerator mass spectrometry the group determined that the first 15 chambers contained only background levels of carbon 14; the levels rose in chambers grown later. By comparing these data with those obtained from the coral, the investigators concluded that the nautilus hatched in 1957 with seven chambers and added eight more in the course of the next year. Over the next decade the animal formed new chambers at a diminishing rate; the last fully sealed chamber took seven months to form and was complet-

By analyzing radiation in bits of shell, workers learn how rapidly the elusive nautilus grows

SPIRALING CHAMBERS of a nautilus are laid bare in a cross section. New findings suggest that the nautilus takes from three to seven months to grow a new chamber.

ed in 1968, a year before the animal was netted.

Paleontologists are eager to know more about the elusive nautilus, because it is the last surviving species of a once prolific class of animals known as shelled cephalopods. During the Mesozoic era, from 200 to 65 million years ago, the oceans teemed with these creatures, tentacled like such modern cephalopods as squid and octopuses but housed in shells that served as both shields and ballast tanks. —*J.H.*

MEDICINE

The 57th Variety

A single amino acid variati·n may contribute to diabetes

Two genetically based studies have helped to advance the fight against diabetes. One investigation has yielded new insight into the cause of insulin-dependent diabetes; the other breaks ground in an approach that may have long-range potential to cure the disease. Insulin-dependent diabetes, thought to be an autoimmune disorder, arises when at least 90 percent of the insulin-producing beta cells of the pancreas are destroyed. The consequent lack of insulin results in elevated blood-glucose levels.

Hugh O. McDevitt, John A. Todd and John I. Bell of the Stanford University School of Medicine have discovered that the replacement of a single amino acid in certain proteins encoded by HLA (human leukocyte antigen) genes greatly increases a person's risk of developing diabetes. HLA proteins are carried on the cell surface and help to initiate the immune response.

The HLA genes are grouped in a number of distinct regions. McDevitt and his colleagues focused on the HLA-D region because about 95 percent of insulin-dependent diabetics possess one or both of the HLA-D alleles (variant forms of a gene) DR3 and DR4. Recent discoveries have shown that HLA-DQ alleles, each of which tends to occur together with a specific DR variant, are also strongly associated with insulin-dependent diabetes and may actually have a more important connection to the disease. (An individual inherits one DR and one DQ allele from the mother and another set from the father; each of the four resulting proteins consists of an alpha chain and a beta chain.)

The apparent significance of the DQ genes convinced the Stanford team to try to pinpoint the amino acid sequences that might play a role in diabetes. They report in *Nature* that they began by comparing the amino acids encoded by two DQ variants that are often inherited with the DR4 gene: DQ*w*3.2, which is very common in diabetics, and DQ*w*3.1, which is not. They found that the genes differed at only four sites, all of which specified amino acids found on the beta chain.

Further analyses homed in on the diabetes-related site. It turns out that the DQ*w*3.2 allele—as well as all other DQ variants that have been positively associated with insulin-dependent diabetes—encodes the amino acid valine, serine or alanine, but not aspartic acid, at position 57 on the beta chain. In contrast, all the DQ alleles that are poorly represented in diabetics code for aspartic acid at that position.

In addition the workers found that each of 39 randomly selected diabetics lacked the position-57 aspartic acid code in at least one of their DQ alleles and that 90 percent had inherited Asp-57-negative DQ alleles from both parents. The investigators conclude that having at least one DQ gene specifying aspartic acid at position 57 generally makes a person resistant to insulin-dependent diabetes; the presence of two such genes confers almost complete protection, and the absence of Asp-57 in the DQ alleles from both parents results in increased susceptibility.

How might a simple amino acid substitution in an individual's HLA-DQ proteins help to cause diabetes? Although the DQ protein presumably is not the only genetically controlled factor that contributes to the disease, Asp-57-negative DQ molecules probably have a three-dimensional structure different from that of other DQ molecules. The difference may affect the virulence of an autoimmune attack on the pancreas.

On the treatment front, Richard F. Selden and his co-workers at the Massachusetts General Hospital recently coaxed the human insulin gene to function in mice. They first inserted a human insulin gene and a genetic marker into a plasmid: a circular piece of bacterial DNA. They then put the plasmid into a culture of fibroblasts, or connective-tissue cells, and cloned the cells that incorporated the insulin gene in their DNA and produced the hormone.

Within a week after the workers injected the genetically altered fibroblasts into the abdomen of 10 diabetic mice, the blood-glucose levels of five of the animals began to fall. Because insulin removes glucose from the circulation, the decline indicated that the genes were in fact producing functional insulin. The levels continued to fall for about three weeks, at which point the mice died—ironically, from too much insulin. Clearly considerable work has yet to be done before cells carrying the insulin gene can be introduced into human diabetics. Nevertheless, the fact that the implanted fibroblasts produced insulin is encouraging. —*Ricki Rusting*

Alzheimer's Proteins

Is the cause too much amyloid protein or too much A68?

Alzheimer's disease runs a grim course: its victims suffer from progressive and incapacitating senility and eventually die from the disorder. More than two million Americans are currently thought to have Alzheimer's and 100,000 die from it annually. Two proteins, A68 and the amyloid beta protein, have been implicated in the disease. Recently A68 has replaced amyloid as the prime suspect in causing the disorder.

Both proteins are found in only two groups of adults: those with Alzheimer's disease and those over the age of 35 who have Down syndrome. In the brain of these individuals the amyloid beta protein forms plaques; the A68 protein is found in neurofibrillary tangles. Both plaques and tangles contribute to the death and destruction of brain tissue, a process expressed in the patient as dementia.

The amyloid beta protein was the first to emerge as a key factor in the disease. In 1984 George G. Glenner and his colleagues at the University of California at San Diego Medical Center demonstrated that the amyloid plaques found in the brain of Alzheimer's victims are identical with those found in older Down syndrome patients, and shortly thereafter several investigators announced that the gene for amyloid beta protein is on chromosome 21, the chromosome implicated in Down syndrome. This was cause for much excitement. Because many of the pathologies associated with Down syndrome are believed to result from the overproduction of specific proteins (all Down patients are trisomic for chromosome 21, having three instead of the normal two copies), there was reason to think duplication of the amyloid gene, resulting in excessive production of amyloid, might explain Alzheimer's. In fact, Jean-Maurice Delabar of the Hôpital Necker in Paris reported finding just such a duplication of the amyloid gene in tissue taken from a deceased Alzheimer's patient.

Reports in *Science* contradict that finding. In three separate studies blood and brain samples from more than 100 Alzheimer's victims were tested postmortem for duplication of either the amyloid gene or genes closely associated with it on chromosome 21. In all cases the amount of DNA coding for the amyloid gene in the Alzheimer's group was almost identical with the amount in the control group (individuals who showed no signs of Alzheimer's) and was two-thirds of the amount measured in a group of Down syndrome patients. "If duplication of that gene does occur," says Peter H. St. George-Hyslop of the Massachusetts General Hospital and an author of one of the studies, "it does so very infrequently, and therefore cannot be closely associated with the disease."

Now the A68 protein is emerging as a key factor in the onset of Alzheimer's disease. According to Peter Davies of the Albert Einstein College of Medicine of Yeshiva University, A68 is normally found in the brain of all human infants, where it is believed to mediate programmed cell death, a process by which surplus neurons are eliminated from the brain in a highly controlled way. When the process is completed, by the age of two, A68 disappears from the brain. It reappears in the brain of Alzheimer's victims and precedes the onset of disease. Data presented by Davies at the annual meeting of the Society for Neuroscience indicate that A68 is consistently found in the brain of Alzheimer's and Down syndrome patients but is virtually absent from the brain of normal adults, including elderly people.

There appear to be two ways A68 might cause the disease. Either the protein triggers the growth of neurofibrillary tangles or it mediates programmed cell death much as it is believed to do in infants. Beyond this, biologists are asking what factors might activate the A68 gene after many years of nonfunction. Davies believes there may be an environmental component. More important at the moment is determining the chromosomal location of the A68 protein gene. If A68 is mapped to chromosome 21, that will again raise the possibility that Alzheimer's is associated with a genetic defect, but until that is either proved or disproved the link between Down syndrome and Alzheimer's disease will remain a mystery.

David Patterson, president of the Eleanor Roosevelt Institute for Cancer Research in Denver, believes Alzheimer's, like cancer, may have more than one cause. Davies disagrees. "It is very hard for me to believe that Alzheimer's disease is heterogeneous, because it is so incredibly predictable." The location of the A68 gene is expected to be announced within the next two or three months; when that happens, another big step in the understanding of Alzheimer's disease will have been taken. *—Laurie Burnham*

TECHNOLOGY

Heavy Metal

NASA and the Air Force differ over cargo launchers

As the National Aeronautics and Space Administration strives to prepare the space shuttle for its next launch, scheduled for June, constituencies are forming for two candidate heavy-payload launch systems that might supplement the shuttle in the 1990's. The discussions have brought to a boil long-simmering disagreements between NASA and the Air Force, which has been convinced, since the *Challenger* disaster, of its need for a means of access to space independent of NASA.

Both parties point to undertakings that will necessitate putting massive loads in orbit. For example, a heavy-payload launcher could ferry components of the planned space station into orbit, thereby easing the burden on the hard-pressed shuttle fleet; it could also loft planetary probes now scheduled for launches on the shuttle. Future communications and reconnaissance satellites might need heavier launchers; so would a "Star Wars" deployment.

NASA is pushing hard for the development by the mid-1990's of an unmanned, cargo-only version of the space shuttle known as Shuttle-C, which would carry a payload of between 100,000 and 150,000 pounds (more than twice the payload of the manned shuttle) into a low earth orbit. The concept, now under preliminary study by aerospace manufacturers, would retain two or three of the shuttle's main engines, the external tank and the advanced solid-fuel rocket motors that are now being developed to replace the shuttle's current boosters. The shuttle orbiter itself, or most of it, would give way to cargo, which would sit either where the current shuttle orbiter is mounted—a scheme that would allow existing launch facilities to be used—or in front of the external tank. In most versions of Shuttle-C only the solid-rocket motors would be reused.

Rear Adm. Richard H. Truly, the associate administrator of the office of space flight at NASA, told aerospace executives at a conference in November that Shuttle-C will, for heavy payloads, be "the cheapest and most reliable launch system available by the mid-1990's." A two-engine version might cost as little as $1 billion to develop. Advocates such as James R. Thompson, director of the George C. Marshall Space Flight Center (which would have a large role in developing the vehicle), point out that besides taking advantage of existing production lines, the cargo shuttle would have the inherent reliability of a system based on a manned vehicle.

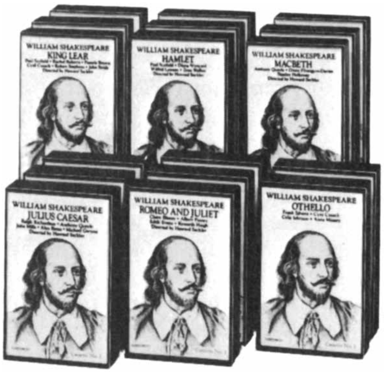
The Air Force opposes the plan, in part because Shuttle-C would have the same high launch costs and low launch rates as the manned shuttle. The Air Force probably also recognizes that if Congress supports Shuttle-C (which is by no means certain in the current budget climate), it is not likely at the same time to fund the Air Force's own and more ambitious Advanced Launch System.

Unlike Shuttle-C, the Advanced Launch System (ALS) would be designed from scratch as a cheap and reliable cargo ferry. This "blue-collar cargo transport system," as Col. John R. Wormington of the Air Force describes it, could lift payloads comparable to those of Shuttle-C. According to the Air Force, the ALS is intended to reduce launch costs by a factor of 10,

Permanent pleasures.

Choose one of these fine works and save up to $305⁵⁰

You simply agree to buy 4 additional books within the next two years.



Shakespeare: 6 Great Tragedies

for **$19.95** (List prices total $179.60)

Book-of-the-Month is proud to present 6 of Shakespeare's greatest tragedies: *Hamlet, King Lear, Macbeth, Othello, Romeo & Juliet* and *Julius Caesar.* Paul Scofield, Ralph Richardson, Albert Finney, Dame Edith Evans, Claire Bloom, John Mills and many more of England's most gifted Shakespearean actors perform memorably in complete readings of each play. 17 cassettes in all.

The Compact Edition of The Oxford English Dictionary

for **$27.95** (Pub. price $195)

"The most complete, most scholarly dictionary of the English language"—*The Christian Science Monitor.* Through photoreduction, the original 13-volume set has been reproduced in this two-volume *Compact Edition.* A Bausch & Lomb magnifying glass is included.

Julia Child: The Way to Cook 6 One-hour Videocassettes

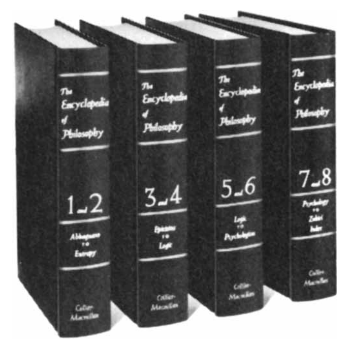
for **$39.95** **(List prices total $179.70)**

America's favorite cook demonstrates the basic techniques of good cooking with step-by-step clarity in her entertaining style. In this 6-hour videocassette set—Meat; Poultry; Fish & Eggs; Vegetables; Soups, Salads & Breads; First Courses & Desserts—she provides professional tips on how to master everything from thickening sauces to butterflying a chicken. Six recipe booklets are included. Available on VHS or Beta.

The Story of Civilization by Will and Ariel Durant

for **$29.95** (Pub. prices total $335.45)

For almost half a century Will and Ariel Durant traced the continuity of world history—the religions and philosophies, the political and economic tides, the arts and sciences, the customs and conquests—to show the foundations of society today. A Book-of-the-Month Club exclusive for almost 50 years, the Durants' illustrated masterwork is history come alive.



The Encyclopedia of Philosophy

for **$24.95** (Pub. price $225)

Regarded as the most comprehensive encyclopedia of philosophy ever published, this superb set—compiled with the world's foremost philosophers—encompasses all aspects of ancient, medieval, modern, Eastern and Western thought. The four volumes represent an essential aid for students and a rewarding reference source.

Facts About Membership. As a member you will receive the *Book-of-the-Month Club News®* 15 times a year (about every 3½ weeks). Every issue reviews a Selection and more than 100 other books, which are carefully chosen by our editors. If you want the Selection, do nothing. It will be shipped to you automatically. If you want one or more other books—or no book at all—indicate your decision on the Reply Form and return it by the specified date. *Return Privilege*: If the *News* is delayed and you receive the Selection without having had 10 days to notify us, you may return it for credit at our expense. *Cancellations*: Membership may be discontinued, either by you or by the Club, at any time after you have bought 4 additional books. Join today. With savings and choices like these, Book-of-the-Month Club is where book lovers belong.

BOOK-OF-THE-MONTH CLUB®

Book-of-the-Month Club, Inc., P.O. Box 8803, Camp Hill, PA 17011-8803 **A170-1**

Check one box only.

- ☐ 923. Shakespeare (17 cassettes) **$19.95**
- ☐ 905. Compact OED **$27.95**
- ☐ 928. Way to Cook (Beta) **$39.95**
- ☐ 929. Way to Cook (VHS) **$39.95**
- ☐ 913. The Story of Civ. **$29.95**
- ☐ 917. Ency. of Philosophy $24.95

Please enroll me as a member of Book-of-the-Month Club and send me the item I have checked at left, billing me for the appropriate amount, plus shipping and handling charges. I agree to buy 4 more books during the next two years. A shipping and handling charge is added to each shipment.

Name____________________ 7-64
(Please print plainly)

Address____________________ Apt._____

City____________________

State____________________ Zip________

Prices shown are U.S. prices. Outside the U.S. prices are generally higher. All orders are subject to approval.

to about $400 per pound—an objective independent experts regard with amused skepticism.

In most of the conceptions now being studied at aerospace companies the ALS would be built around one or several engines fueled with liquid hydrogen and liquid oxygen. The Air Force envisions an interim ALS, in service as early as the mid-1990's, that would probably use additional strap-on solid rockets. (Admiral Truly, in contrast, sees no need for any system based on new technology before the late 1990's.) In a completed ALS, which would be ready by the late 1990's, the expendable strap-on boosters would probably be replaced by a booster vehicle that would fly back to the earth for reuse. In one radical proposal, however, the ALS would combine 200 copies of a currently available rocket motor in an enormous cluster.

This January NASA and the Air Force will deliver joint studies on possible heavy launch systems to Congress. Both Shuttle-C and the ALS face considerable skepticism. Congressional staff point out that Shuttle-C is not absolutely necessary for the space station, and that an interim ALS is hard to justify unless a Star Wars defense is to be built in the 1990's. That circumstance in itself could doom the interim ALS politically, because Congress has forbidden expenditures intended to facilitate an early deployment of a Star Wars system. —*T.M.B.*

Flat Out

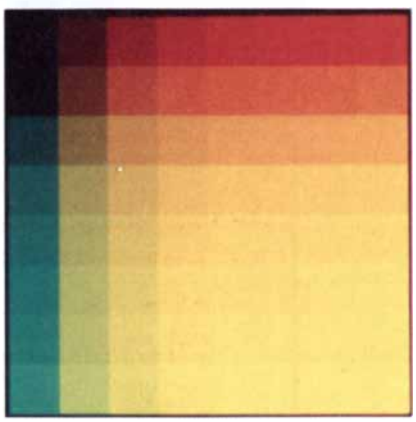
Japan scrambles to dominate market for flat video screens

Flat screens, the long-heralded replacement for the bulky cathode-ray tubes of televisions and computer displays, are coming into their own. Tiny color televisions with flat-panel displays are on the market, and one Japanese maker recently demonstrated color displays measuring 14 inches on a diagonal. William J. Caffery of the Gartner Group, Inc., a computer consulting company in Stamford, Conn., believes flat-panel displays are "near the takeoff point." As in many consumer technologies, Japan appears to be calling the shots.

One flat-screen technology, the plasma display, produces a bright red light by ionizing a gas. Monochromatic plasma displays are found in some portable computers, and Japanese manufacturers have demonstrated color displays, which produce ultraviolet light that stimulates a layer of phosphors to give off visible light in various colors. Crystalline materials that emit light when they are energized are the basis of yellow "electroluminescent" computer displays; color versions are also being developed. Both plasma displays and electroluminescent screens have traditionally been handicapped by their high power consumption.

Improvements to liquid-crystal displays (LCD's) are driving most of the progress in flat screens. In an LCD a thin layer of an organic chemical whose cylindrical molecules tend to align themselves in one direction, in a kind of crystalline order, is trapped between two sheets of glass, which in turn are sandwiched between two crossed polarizing filters. The dominant orientation of the molecules is twisted by 90 degrees from one side of the liquid-crystal layer to the other, so that the layer ordinarily rotates the polarized light admitted by one of the filters, enabling it to pass through the other.

One of the glass sheets enclosing the liquid crystal is ruled with thin, vertical electrodes and the other with horizontal electrodes. Applying a voltage across any pair of vertical and horizontal electrodes activates the pixel, or picture element, at their intersection, causing the molecules there to become reoriented, so that they no longer rotate polarized light. Light is now blocked by crossed polarizers: the pixel seems to change from light to dark. Color can be added by means of color filters.



THIN-FILM electroluminescent display developed by Planar Systems, Inc., of Beaverton, Ore., actively emits light from a crystalline material. The flat screen measures about six inches diagonally and is about three-quarters of an inch thick.

The most primitive LCD's—the black-on-gray or gray-on-green displays common in laptop computers and digital watches—have a restricted viewing angle; in larger screens contrast, which is low to begin with, is further reduced because some of the electric charge intended for a particular pixel activates extraneous pixels. A new "supertwist" liquid crystal sustains a 270-degree twist in molecular orientation, thereby improving contrast.

Flicker is another problem that appears as screen size increases: only one row of pixels can be addressed at a time, and the molecules tend to return to their resting position when charge is absent. "Ferroelectric" liquid crystals are the subject of intensive research because the characteristics of their molecules result in a molecular arrangement that has "memory," a property that could prevent flicker. Such molecules are also less affected by leakage of current, which helps the screen to achieve high contrast.

The "active matrix" LCD addresses the contrast problem with an advance in the electronics controlling the display. In a typical active matrix a thin-film transistor built into the screen controls each pixel, switching the pixel fully on or off in spite of current leakage. Low manufacturing yield daunts efforts to produce such screens in large sizes, however. Every transistor—and a full-size television display would contain hundreds of thousands of them—must work perfectly. The biggest active-matrix panels now available commercially, such as the one in a portable color television made by the Sharp Corporation, measure about three inches on the diagonal and are expensive.

U.S. companies working on active matrixes are focusing chiefly on military applications, such as fighter-plane cockpits, where resolution, visibility and weight are at a premium and price is not. A million-pixel six-inch display has been demonstrated by the General Electric Company, and Ovonic Imaging Systems, Inc., is at work on an eight-inch active matrix. Characteristically Japan has targeted the consumer, Caffery says, applying revenue from sales of portable flat-panel televisions to developing ways of making large displays cheaply. Even so, according to John A. van Raalte of the Society for Information Display, replacing the cathode-ray tube in living-room televisions with a flat screen might take another 15 years. —*T.M.B.*

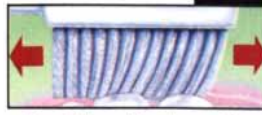
SHOP BY MAIL **THE LIFESTYLE RESOURCE™** SHOP BY PHONE

At The Lifestyle Resource we give you all the facts and details necessary to make an informed purchase. But your satisfaction is our primary concern. If your purchase doesn't meet your expectations, just return it within 30 days for a prompt refund.

THE NEW INFRALUX

Pain emanating from inflamed joints and muscles can be excruciating. Just ask sufferers of bursitis, neuralgia, or arthritis. Infrared heat is the natural remedy preferred by doctors and therapists to dispel such nerve-racking pain. Recently redesigned, the Infralux infrared therapeutic heating unit for personal use now is contoured, angled for reaching back and neck. The heating surface is tripled and the replaceable 10-watt bulb gives 45% more heat. Hi/low settings. Plugs into household outlet. UL listed. Unit with case **$39.95** #1140.

YOUR SMILE IS BEAUTIFUL

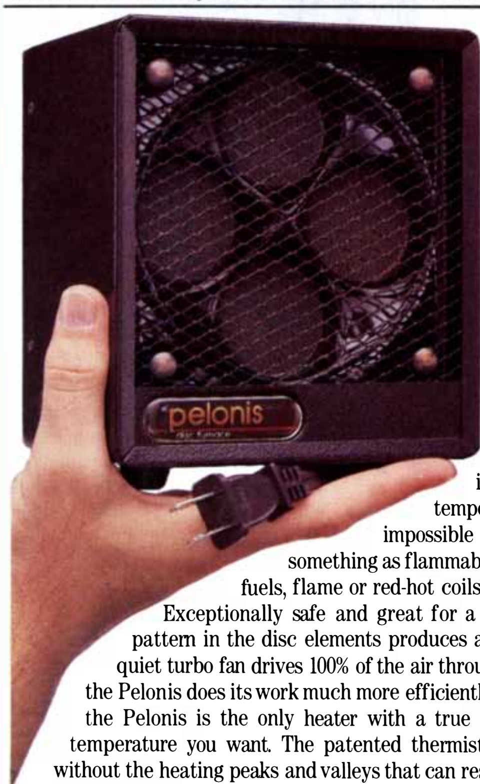


Hand brushing's motion misses plaque.

The counter-rotational movement cleans plaque.

But you're only fooling yourself if you think it's proof of healthy teeth and gums. The unseen enemy is a natural mouth substance deposited on the teeth - plaque. Unchecked, plaque develops into rock-hard tartar behind which decay breaks down the tooth enamel and invades the soft, sensitive substructure. You have to remove the plaque before it hardens to prevent the process. Until recently there was no easy solution to this problem, but now there's Interplak, a precision-engineered, university-tested, three-speed plaque removal instrument. Its 10 independently driven bristle tufts rotate around their own axes: each tuft spins in the direction opposite to the spin of its neighbor at 4,200 rpm. The bristles are soft and rounded, adhering to dentistry recommendations, so that virtually every particle of plaque—*98% of it*—is removed gently and without irritation. Interplak even reaches between teeth and around orthodontic braces. Children think it's fun, so they're eager to brush. Interplak is cordless and easily stores in a free-standing or wall-mounted recharger/holder. Individual color-coded brush heads are available. The Interplak Home Plaque Removal Instrument with charger and two brush heads. **$99.95** #1110. Two additional brush heads **$29.95** #1111.

THE SIX INCH, SIX POUND FURNACE

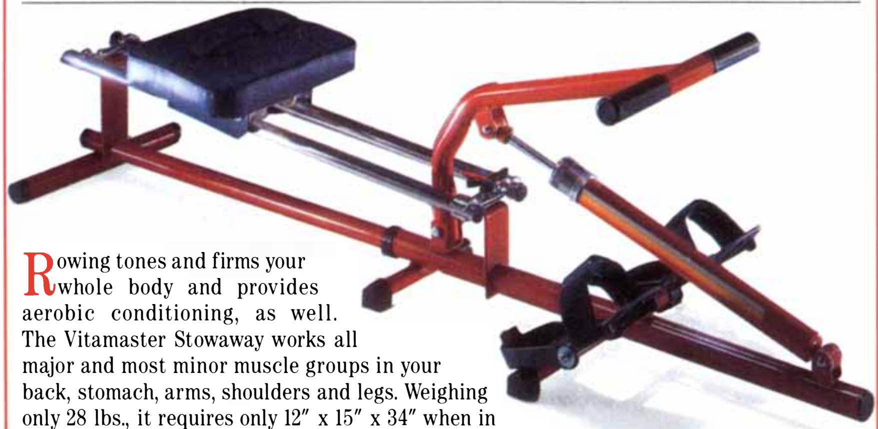


Imagine heating a 17′x 20′ room for 24 hours for 68 cents! In tests conducted by the National Certified Testing Laboratories, the Pelonis Disc Furnace raised the temperature of a 17′x 20′ room from 50°F to 70°F and maintained the temperature at that level for 24 hours. The costs were based on the national average for electricity of 7 cents per kilowatt hour. The secret of the revolutionary Pelonis Disc Furnace is a scientific breakthrough: barium titanate honeycomb disc heating elements. Let's look more closely at some of the unique features of this remarkable product. *Safety:* despite its enormous heat output of 5,200 BTU's, the tiny Pelonis is incredibly safe. Utilizing special ceramic low temperature heating elements, it is virtually impossible for the heating elements to ignite even something as flammable as tissue paper. There are no combustible fuels, flame or red-hot coils and a safety grille gives added protection. Exceptionally safe and great for a child's room. *Efficiency:* the honeycomb pattern in the disc elements produces an extremely large heating surface, and the quiet turbo fan drives 100% of the air through. No other heater can do that. As a result the Pelonis does its work much more efficiently than an ordinary electric heater. *Comfort:* the Pelonis is the only heater with a true thermostat that allows you to dial in the temperature you want. The patented thermistor circuitry then holds that temperature, without the heating peaks and valleys that can result from cycling. *Durability:* the Pelonis is built like a tank, with a heavy-gauge metal case, industrial-grade components, and a meticulous attention to detail. It carries a 5-year limited warranty. *Versatility:* this six pound, six inch cube will go anywhere. The Pelonis allows you to lower the heat for the entire house and selectively heat only those areas you occupy. The Pelonis Disc Furnace was introduced in Canada (where they know a lot about cold) in 1984; last year it outsold all other electric heaters combined! We are pleased to introduce this latest technology in electronic heating. Includes a built-in handle and washable air filter. Standard 110 volts AC, UL listed. **$189.95** #1811.

ZONE OF CALM

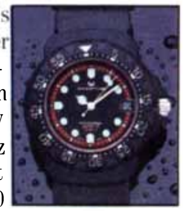
Noise produces stress in the human organism. Today the volume of civilization nearly everywhere seriously interferes with our abilities to relax, read, sleep, concentrate or otherwise function at optimum efficiency. But you needn't be victimized by noise pollution. The Marsona Sound Conditioner electrically synthesizes a variety of pleasing natural sounds that help mask out and reduce the annoyance of unwanted noise. You can simulate the sounds of ocean surf, summer rain, mountain waterfalls. You control not only the volume but also wave pattern, wave or rain rhythm, the seeming nearness or distance of the source. The Marsona is the finest instrument of its kind on the market, solid-state electronics designed for continuous use (electrical draw only 5 watts). UL-listed. **$139.95** #1860.

ROWING FOR FITNESS



Rowing tones and firms your whole body and provides aerobic conditioning, as well. The Vitamaster Stowaway works all major and most minor muscle groups in your back, stomach, arms, shoulders and legs. Weighing only 28 lbs., it requires only 12″ x 15″ x 34″ when in operation and collapses after use for compact storage. Actually takes up less room than a suitcase. The Stowaway is constructed of heavy tubular steel and has a full-size 54″ bed. The contour seat is padded for comfort and mounted on nylon rollers which glide on chrome rails. Single rowing arm with hydraulic tension resistance has 12 adjustable tension settings—from light to heavy—to provide a varied workout. Molded plastic pedals with adjustable footstraps rotate individually to simulate rowing action. Non-mar footpads protect floor. Comes assembled. Save yourself an expensive gym membership. For the one-time cost of **$119.95** you can exercise in the privacy of your own home. #1970.

WHICH WATCH?



You choose. You get this year's most sought-after chrono style and a totally professional competition sports watch, powered by a precision Citizen quartz movement. Dive with it: it can go an incredible 660 feet deep. Monitor elapsed time for running, swimming, or cycling: the deep-knurled, uni-directional bezel can't accidentally reverse, and you can set it with gloves on. The case is fiberglass bonded to stainless steel; scratch resistant mineral glass crystal seals out dust and water. Shock resistant 3 hand movement with date read out window. Luminous hands let the action go on even in the dark. You've seen watches just like this for $145. Now, enjoy the same exceptional quality and look for a fraction of the cost. Adjustable PVC band can be trimmed for perfect fit. Yellow with black, #1940; Black with red, #1950; Black with blue, #1960. **$59.95.**

NEW NEON

Nuon Klocks' glow can enliven any room at home or the office. Light up a kitchen, den, bar or put one in a child's room as a combination clock/night light, or just indulge in a little Art Deco nostalgia.

Now, you can have the pizzazz of neon without its fragility or its unbelievable cost. The Nuon Klocks innovative use of PVC tubing impregnated with brillant colored fluorescent dyes, allow the sculpted numerals and hands to take on a vibrant neon-like appearence when illuminated by an ultra-violet light hidden in the clock base. Nuon Klocks' glow will enliven any room at home or the office. Light up a kitchen, den, bar or put one in a childs room as a combination clock/night light, or just indulge in a little Art Deco nostalgia. A single AA battery powers the accurate quartz movement for about a year. Plug into any wall outlet to power the long-life (8,000 hr.) UV tube with on/off switch. Modern, wraparound design in sleek acrylic allows maximum visibility. Klocks are free-standing or can be wall-mounted. 9 ½″ H x 13″ W x 3 ¼″. D. Two models: "Big Time" #1920, or "[illegible] and Sunset" #1930. **$59.95.**

GIVE YOUR BACK A BREAK

Lower back pain is one of the commonest of human ailments. If you suffer from it, or if you want to avoid it, *you need this*—The Back Machine. Developed in cooperation with leading orthopedic physicians, and in wide use by hospitals and physical therapists, this remarkable 12″ x 17″ device—no bigger than a cushion—compensates for stress, fatigue, bad posture or poorly designed chairs. The Back Machine adjusts to nine positions to conform to your own unique spinal geometry: just unzip the washable cover and rotate the adjustable roller until the lumbar curve is precisely where you want it. Perfect for home, office, the car or plane (especially those long trips) it's light weight and portable. Durable steel and plastic construction, wi[illegible] velcro straps to attac[illegible] any chair. **$39.95,** [illegible].

SCIENTIFIC AMERICAN January 1988 Volume 258 Number 1

The Antarctic Ozone Hole

Each spring for the past decade the ozone layer in the atmosphere has thinned at the South Pole. Is the loss an anomaly, or is it a sign that the ultraviolet-absorbing layer is in jeopardy globally?

by Richard S. Stolarski

In 1985 atmospheric scientists of the British Antarctic Survey published a completely unexpected finding: the springtime amounts of ozone in the atmosphere over Halley Bay, Antarctica, had decreased by more than 40 percent between 1977 and 1984. Other groups soon confirmed the report and showed that the region of ozone depletion was actually wider than the continent and extended roughly from 12 to 24 kilometers in altitude, spanning much of the lower stratosphere. There was, in essence, an ozone "hole" in the polar atmosphere.

The discovery disturbed scientists and the public alike because it suggested that the stratospheric layer of ozone surrounding the globe might be in greater jeopardy than atmospheric models had predicted. A precipitous erosion of the layer would be cause for much concern. Ozone, a three-atom molecule of oxygen, constitutes less than one part per million of the gases in the atmosphere, yet it absorbs most of the ultraviolet rays from the sun, preventing them from reaching the earth. Such radiation is sufficiently energetic to break apart important biological molecules, including DNA. It can increase the incidence of skin cancer, cataracts and immune deficiencies and may also harm crops and aquatic ecosystems.

Because these effects are so serious, many investigators, including my colleagues and me at the National Aeronautics and Space Administration, have been racing to determine the causes of the hole, which develops each southern spring within the polar vortex: an isolated air mass that circulates around the South Pole during a large part of the year. (The amount of ozone in the vortex decreases in late August and early September, stabilizes in October and increases again in November.) Until we know why the hole forms, we will not be able to determine whether it has global implications or will remain confined to the Antarctic stratosphere, where meteorological conditions are unique.

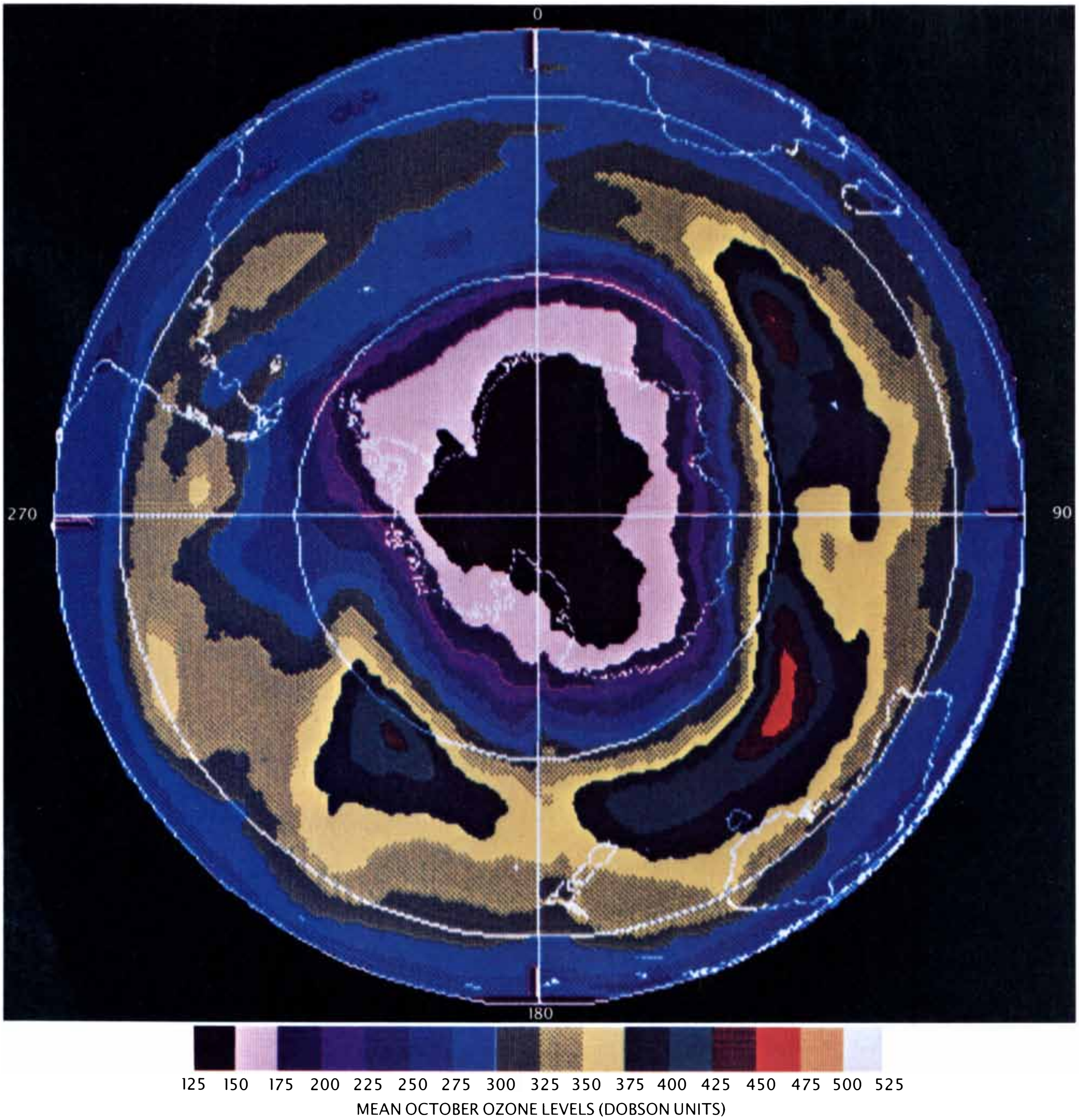
The research effort has taken advantage of several measuring tools, some based on the ground and some carried by balloons and by satellites. The balloon-borne instruments generally examine the chemistry of the air through which they travel. The ground- and satellite-based equipment perform remote observations, such as measuring the thickness of the layer, or column, that would result if all the ozone directly above an observer at the surface of the earth were brought to standard temperature and pressure. To obtain the thickness, which is generally recorded in Dobson units, or hundredths of a millimeter, investigators measure the radiation striking the earth at slightly different wavelengths, some of which are strongly absorbed by ozone and some not. (Satellite-borne instruments record reflected light.) If the amount of radiation at the absorbed wavelengths increases with respect to that at the unabsorbed wavelengths, the amount of ozone has decreased; conversely, if the radiation at the absorbed wavelengths declines, ozone has increased.

The investigations have drawn on international resources. In 1987, for instance, some 150 scientists and support personnel representing 19 organizations and four nations met in Punta Arenas, Chile, to conduct the most ambitious study yet, the Airborne Antarctic Ozone Experiment. That experiment, which determined that the ozone hole was at its largest in 1987, involved not only ground- and satellite-based measuring devices and balloons but also airborne laboratories. A retrofitted DC-8 passenger plane and an upgraded version of the U-2, the ER-2, flew into the depleted region several times to gather detailed information about its size and chemistry.

The 1987 campaign, like other recent studies, focused on two of the foremost explanations for the ozone hole. One theory assumes pollutants are a cause; the other emphasizes a natural shift in the air movements that transport ozone-rich air into the polar stratosphere during the southern spring.

Concern about pollutants actually predated any evidence of the damage they might cause. In 1971, when supersonic transport planes were expected to become a major presence in the skies, many investigators worried that the resulting emission of water vapor and nitrogen oxides (NO_x) would have a harmful effect on the atmosphere at high altitudes. Laboratory studies had shown that both types of gas could attack ozone. The proposed supersonic fleet never materialized, but in the years that followed, rising levels of nitrous oxide (N_2O) in the environment from increased combustion and increased use of nitrogen-rich fertilizers led to similar concerns. These paled in 1974 when Mario J. Molina and F. Sherwood Rowland of the University of California at Irvine sounded an alarm over the increasing use of compounds known as chlorofluorocarbons (CFC's). Since that time the so-called fluorocarbon problem has dominated a great deal of the research in the field.

Chlorofluorocarbons, as the name suggests, consist of chlorine, fluorine and carbon. First introduced some 60 years ago, these gases have served as coolants for refrigerators and air conditioners, propellants for aerosol sprays, agents for producing foam, and cleansers for electronic parts. The compounds, which continue to have wide application, were once considered to be virtually ideal industrial chemicals because they are highly stable and unreactive and therefore nontoxic. Ironically it is the inertness of the compounds that makes them potentially troublesome



MAP OF OZONE LEVELS in the atmosphere of the Southern Hemisphere on October 5, 1987, highlights the springtime ozone "hole" (*black, pink, purple*) over Antarctica. In the hole the abundance of ozone is about half that of a decade ago, which averaged about 300 Dobson units. Each unit is a hundredth of a millimeter and refers to the thickness of the layer that would result if the ozone in a slice of the atmosphere were collected at standard temperature and pressure. The map is based on data from the Total Ozone Mapping Spectrometer (TOMS) on board the National Aeronautics and Space Administration's *Nimbus 7* satellite.

for the ozone in the stratosphere.

Inert gases do not degrade readily in the troposphere, the part of the atmosphere ranging from the surface of the earth to an altitude of about 10 kilometers. As a result such substances eventually find their way into the stratosphere, which extends to about 50 kilometers. When molecules rise above some 25 kilometers, roughly where the concentration of ozone peaks, they become subject to the intense ultraviolet light that is absorbed by ozone at lower altitudes. This radiation can break normally stable molecules (such as the chlorofluorocarbons) into more reactive forms (such as chlorine atoms).

Laboratory studies had shown that chlorine readily destroys ozone. Because millions of tons of chlorofluorocarbons were being released into the environment, many investigators argued that continued release would eventually cause the compounds to accumulate in the stratosphere to a level capable of seriously eroding the ozone shield. Moreover, it was likely the destructive process would continue well into the next century—even if chlorofluorocarbon release were to cease immediately—because the chemicals remain in the atmosphere for decades. The two major varieties, numbers 11 ($CFCl_3$) and 12 (CF_2Cl_2), persist for approximately 75 and 100 years respectively.

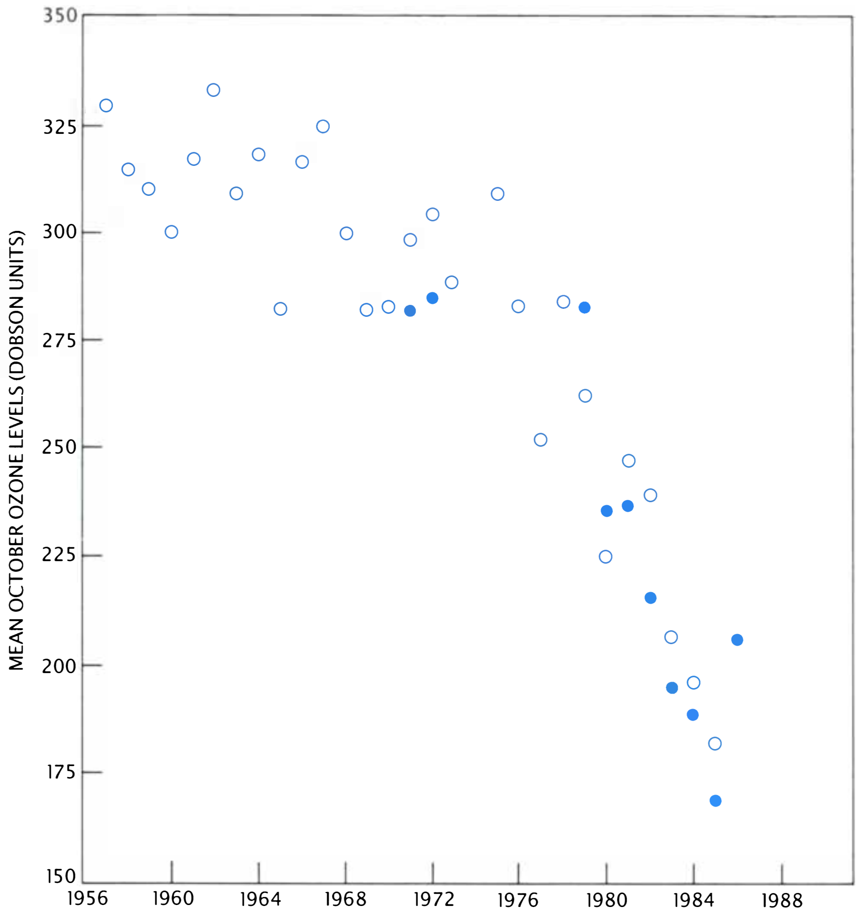
On the strength of such arguments, in 1978 the U.S. banned the use of chlorofluorocarbons in aerosol products, such as certain deodorants and hair sprays. But efforts to control other applications met with little success, in part because of an increasing awareness of the complexity of the chemistry of the stratosphere. For instance, although it was already known that such substances as oxides of nitrogen and hydrogen could by themselves destroy ozone, new calculations showed that nitrogen oxides could interact with chlorine in ways that would actually impair the chlorine's ability to attack ozone.

Then came the British announcement, with its report that the October levels of ozone above their research station had dropped from a normal value of about 300 Dobson units in the early 1970's to about 180 units in 1984. The finding revived public concern about the global ozone layer. At the same time, because of increases in the release rates of the chlorofluorocarbons, policymakers were engaged in ongoing discussions about placing international controls on the compounds. Such discussions led 23 nations (including the U.S.) this past September to sign an agreement to reduce consumption of the substances. The accord, which must be ratified by at least 11 nations before it becomes official at the start of 1989, requires developed nations to freeze consumption at 1986 levels by mid-1990 and to halve usage by 1999.

The chemical phenomenon at the root of the chlorofluorocarbon theory is the ability of small amounts of chlorine to destroy ozone in quantity. A molecule of ozone (O_3) is created when ultraviolet light strikes an oxygen molecule (O_2). A photon splits the molecule into two highly reactive oxygen atoms (O). These quickly combine with intact oxygen molecules to form ozone (O_3). The gas readily absorbs ultraviolet light and dissociates into its component parts (O_2 and O); the freed oxygen atom subsequently joins with another oxgen molecule, thereby re-forming ozone. The gas goes on to dissociate and re-form many times in this way until it finally collides with a free atom of oxygen, forming two stable oxygen molecules [*see top illustration on page 34*]. Given constant conditions, the net result is that ozone settles into a dynamic steady state, in which the rate of its formation is equal to the rate of its removal.

Chlorine shifts this balance and reduces the amount of ozone in the stratosphere by hastening its conversion into two oxygen molecules. More important, the chlorine (like oxides of nitrogen and hydrogen) acts catalytically: it is unchanged in the process. Consequently each chlorine atom can destroy as many as 100,000 ozone molecules before it is inactivated or eventually returned to the troposphere, where precipitation as well as other processes removes it from the atmosphere.

The chemistry thought to drive such destruction is fairly clear-cut. When a chlorine atom (Cl) collides with an ozone molecule, the chlorine steals the ozone's third oxygen atom, forming a chlorine monoxide radical (ClO) and an oxygen molecule. Radicals, which are molecules that have



DECLINE (*left*) in springtime ozone levels over Antarctica was first identified by Joseph C. Farman and his colleagues at the British Antarctic Survey, who have monitored ozone levels directly over Halley Bay since 1956 (*open circles*). After the workers published

an odd number of electrons, are quite reactive. When the chlorine monoxide meets a free oxygen atom—an important step in the catalytic cycle—the oxygen in the chlorine monoxide becomes highly attracted to the free atom and breaks away to form a new oxygen molecule. The "abandoned" chlorine is free to begin ozone destruction anew.

The chlorine catalytic cycle generally does not operate unfettered. Two major kinds of reaction are believed to interfere with ozone destruction, at least in the mid-latitudes. In one case chlorine monoxide reacts with nitric oxide (NO). The oxygen atom in the chlorine monoxide is transferred to the nitric oxide, producing a free chlorine atom and nitrogen dioxide (NO_2). When the latter absorbs visible light, it releases an oxygen atom, which is then available to regenerate ozone [*see bottom illustration on next page*]. The net result is no change in the ozone level.

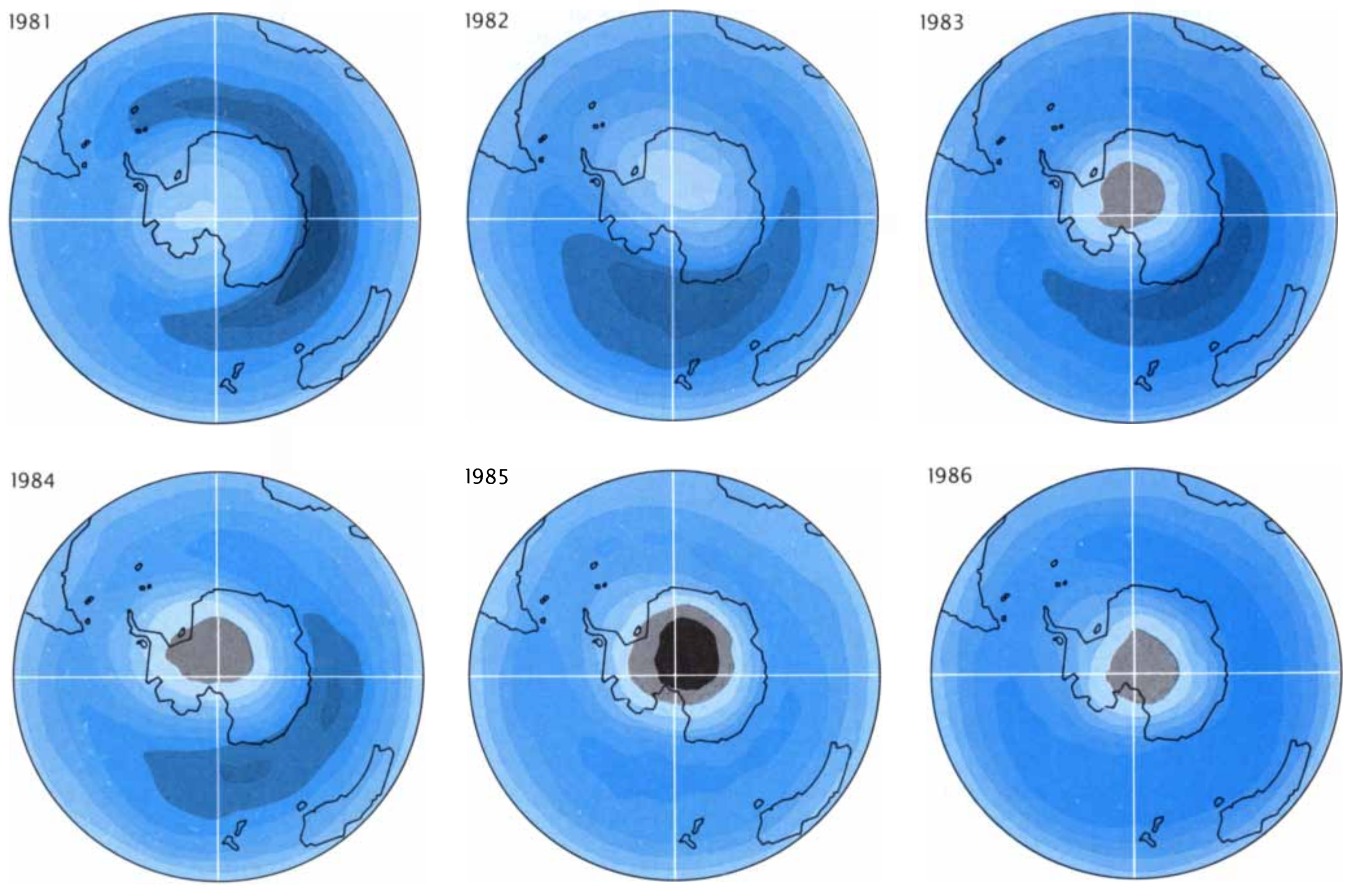
In a second and more important case a chlorine atom or a chlorine monoxide radical combines with another molecule to form a stable product that temporarily acts as a chlorine "reservoir"; when chlorine is thus engaged (as it is most of the time in the normal atmosphere), it is unavailable to attack ozone. Two important reservoirs are chlorine nitrate ($ClONO_2$), formed by the combination of chlorine monoxide and nitrogen dioxide (NO_2), and hydrochloric acid (HCl), formed from a chlorine atom and methane (CH_4). Eventually such reservoirs absorb a photon or react with other chemicals, break apart and free the chlorine to resume its catalytic destruction of ozone.

The existence of these interference reactions has consistently led computer modelers to conclude that chlorofluorocarbons should so far have had a minimal effect on the global ozone layer. The finding that springtime ozone levels at the South Pole have declined by more than 40 percent suggests that if chlorine from chlorofluorocarbons is a cause, then the normal interference reactions are somehow minimized during the Antarctic spring. The question is how.

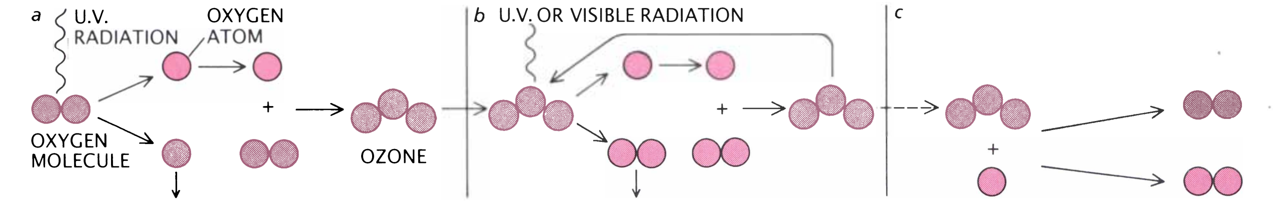
Proponents of the chlorofluorocarbon theory of the ozone hole have identified several different processes that could well minimize the effects of these interferences. For instance, the removal of nitrogen oxides from the stratosphere would facilitate the destruction of ozone. If the oxides were unavailable, they could not combine with chlorine to form the reservoir chlorine nitrate. In addition some process might alter chlorine reservoirs, causing them to release active chlorine, in the form of single atoms or chlorine monoxide, that would destroy ozone.

Many workers suspect that polar stratospheric clouds (PSC's) contribute to such processes. These high-altitude clouds, which are much commoner in the Antarctic region than in the Arctic, form in the winter, when the absence of sunlight and the isolation of the Antarctic region lead to stratospheric temperatures that often dip below −80 degrees Celsius.

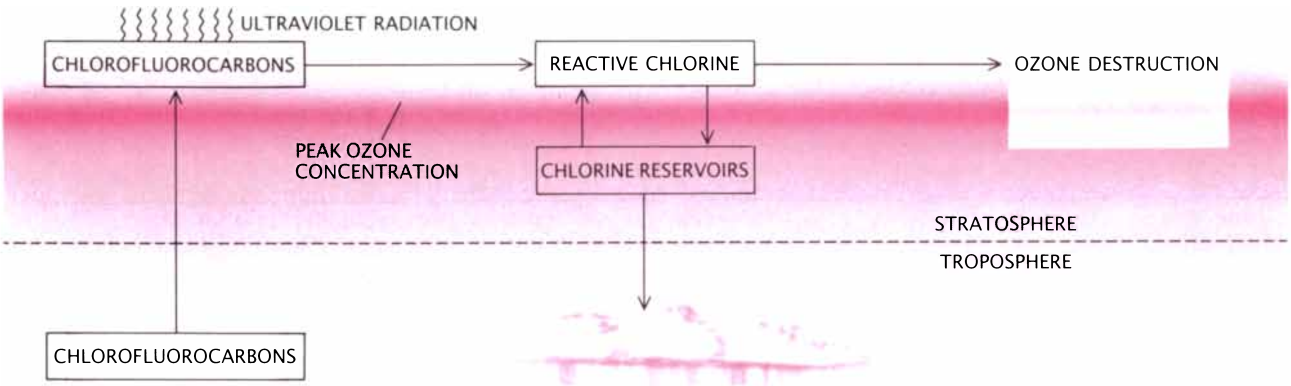
It is conceivable that nitrogen compounds condense and freeze during



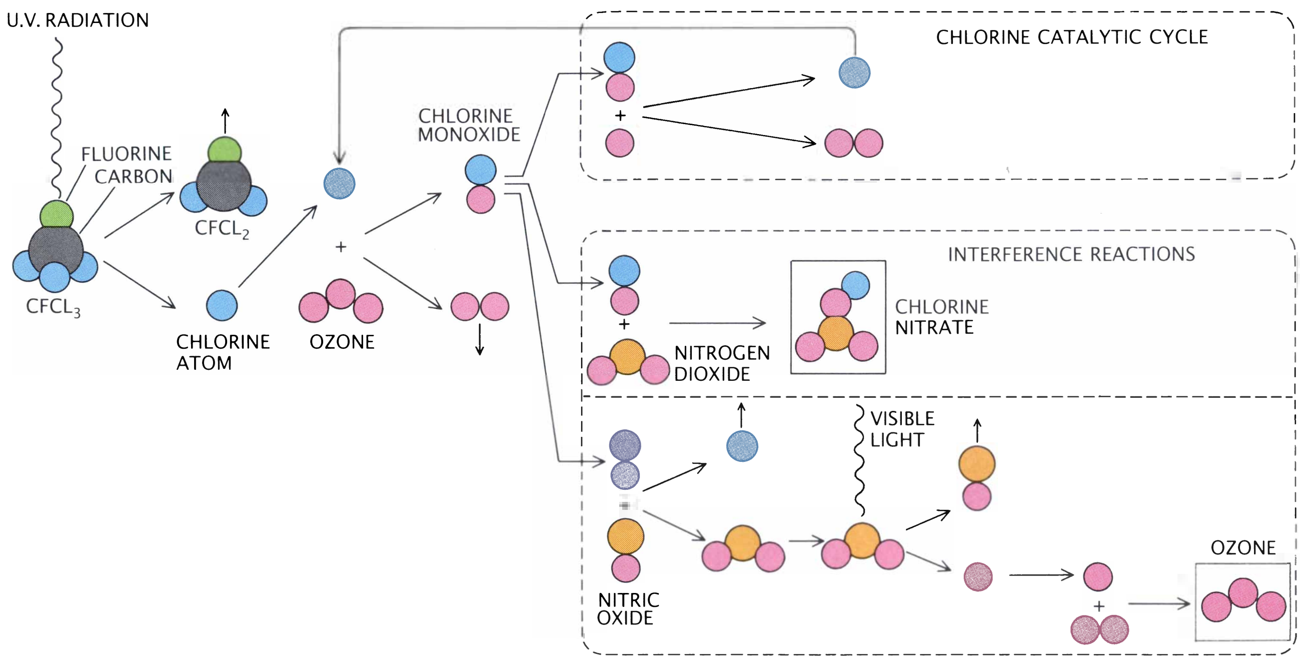
their report in 1985, NASA confirmed it by satellite (*dots*). Other NASA data (*right*) showed that the depleted region is broader than the Antarctic continent and flanked by an ozone-rich region (*crescent*). The maps, which have been redrawn for clarity, are based on TOMS data. Black indicates 150 to 180 Dobson units and is followed in increasing order by gray and deepening blues.



OZONE in the atmosphere absorbs vast amounts of ultraviolet radiation that would otherwise reach the earth. The gas is created (*a*) when a highly energetic ultraviolet photon strikes an oxygen molecule (O_2), freeing its atoms (O) to combine with nearby oxygen molecules. The ozone (O_3) so formed is repeatedly broken apart by photons of ultraviolet or visible light and quickly re-formed, ready to absorb more light (*b*). Ozone "dies" (*c*) when an oxygen atom collides with it, forming two oxygen molecules.



CHLOROFLUOROCARBONS, which are synthetic chemicals, are thought to contribute importantly to the ozone hole. After being released into the troposphere, where they remain inert, the compounds eventually rise into the upper stratosphere, above the region where ozone concentrations (*color*) peak. The ultraviolet radiation there is strong enough to break the molecules apart, freeing chlorine atoms to attack ozone. The destructive effects of the chlorine end when the atoms combine with other substances to form stable chlorine "reservoirs." Such molecules may dissociate in the presence of heat or light, returning chlorine to the stratosphere, but some of them settle into the troposphere, where various processes remove them from the atmosphere.



CHLORINE CHEMISTRY includes processes that promote ozone destruction and processes that hinder it. A chlorine atom can destroy ozone catalytically (*left and top*), without being consumed itself. It first "steals" an oxygen atom from ozone, forming chlorine monoxide (ClO) and a stable oxygen molecule. When the ClO collides with another oxygen atom, the two oxygens combine readily, freeing the chlorine atom to remove another ozone molecule. Other processes interfere with the catalytic cycle. For example, nitrogen dioxide (NO_2) can bind to chlorine monoxide to form a chlorine reservoir (*middle*); while chlorine is thus engaged it cannot react with ozone. Another source of interference (*bottom*) is nitric oxide (NO). It takes the oxygen atom from chlorine monoxide, absorbs visible light and regenerates ozone. Chemically based explanations for the ozone hole suggest that the unique climatic conditions at the South Pole minimize such interferences, allowing the catalytic cycle to proceed efficiently.

the winter, becoming bound up in cloud particles. They would then be unavailable to react with chlorine. At the same time the cloud particles might facilitate the conversion of chlorine reservoirs into active chlorine. Some chemical reactions that occur slowly in a purely gaseous environment may take place significantly faster on the surfaces of particles. In the darkness of the polar winter many chemical processes are essentially at a standstill. Nevertheless, it is possible that the PSC particles trap and slowly modify the major chlorine reservoirs, preparing chlorine monoxide to make a rapid "escape" when the sun begins to shine. Unfortunately the composition of the cloud particles and the exact reactions that take place on their surfaces are still not known.

The chlorofluorocarbon theory of the ozone hole must explain not only how "normal" interferences to the chlorine catalytic cycle are impeded during the Antarctic spring but also how a uniquely polar phenomenon is overcome. In the polar spring the sun is still rather low on the horizon; as a result there is a reduction in the radiation-driven breakdown of ozone molecules and consequently a reduction in the amount of free oxygen atoms available for the chlorine catalytic cycle.

The presence of a reasonable quantity of bromine (Br) in the polar stratosphere could help to compensate for this deficit of free oxygen atoms. The chemical, which is released into the atmosphere from the naturally occurring compound methyl bromide and from fumigants and certain fire extinguishers, can interact with ozone to form a bromine monoxide radical (BrO) and an oxygen molecule. The bromine monoxide in turn can react with chlorine monoxide, forming another oxygen molecule and releasing free chlorine and bromine atoms. (The net result is the transformation of ozone into oxygen.) Such a combined chlorine-bromine catalytic cycle can operate smoothly even when free oxygen atoms are relatively scarce in the environment. Bromine is also efficient at destroying ozone on its own: it initiates a chain of reactions similar to those generated by chlorine but not requiring free oxygen atoms.

Even though many questions remain about the chemistry that might drive ozone destruction in the Antarctic stratosphere, data from a major 1986 study at McMurdo Sound in Antarctica and the preliminary findings of the 1987 Airborne Antarctic Ozone Experiment lend some support to the chlorofluorocarbon theory. For instance, in the springtime the amounts of chlorine monoxide in the ozone hole are elevated compared with the levels at mid-latitudes. Moreover, as predicted, the levels of nitrogen oxides in the hole are severely depressed compared with the levels at the mid-latitudes.

POLAR STRATOSPHERIC CLOUDS over Antarctica form when frigid winter temperatures cause water vapor and possibly other gases, such as nitric acid, to condense and freeze. It has been proposed that the clouds may facilitate the breakdown of chlorine reservoirs, freeing chlorine to destroy ozone when the sunlight returns in the spring.

The data also provide evidence consistent with the suggestion that the reservoirs chlorine nitrate and hydrochloric acid are altered by clouds. The gas-state levels of both reservoirs are low at the beginning of the Antarctic spring (when the hole is forming), but then they rise. The increase suggests that much of the gas was initially stored in an undetectable form, such as in cloud particles. Whether bromine also has an important function, as theory suggests, is not clear. Preliminary findings indicate that its concentration may not be particularly high in the stratosphere over Antarctica.

Support for a chemical explanation of the ozone hole does not eliminate the possibility that natural processes, such as a shift in atmospheric dynamics, are also important. Dynamical processes do not destroy ozone; they simply redistribute it.

The suspicion that a change in dynamics may play a role derives from the fact that the atmosphere is not static. Rather it is a three-dimensional fluid that moves constantly, changing the location and amounts not only of ozone but also of all the chemicals that affect it. If ozone levels, which always fluctuate to some degree, were affected only by the sun, one would expect to find the highest levels in places where the sun is strongest: very high altitudes and very low latitudes. In actuality ozone levels peak not at the top of the stratosphere but in the middle. Moreover, the largest amounts are found not above the Equator, where ozone levels are typically only about 260 Dobson units, but near the poles.

This distribution arises because stratospheric air tends to circulate from high altitudes in the Tropics toward lower altitudes in the polar regions, carrying newly made ozone with it. In the Northern Hemisphere stratospheric air circulates all the way to the North Pole, where the average ozone level reaches about 450 Dobson units in the late winter or early spring. In the Southern Hemisphere the circulation generally goes only as far as about 60 degrees south latitude for most of the year; ozone levels there peak at about 380 Dobson units. Peculiarities of Antarctic meteorology, such as the polar vortex, impede the ozone-rich air from moving farther south until late in the spring.

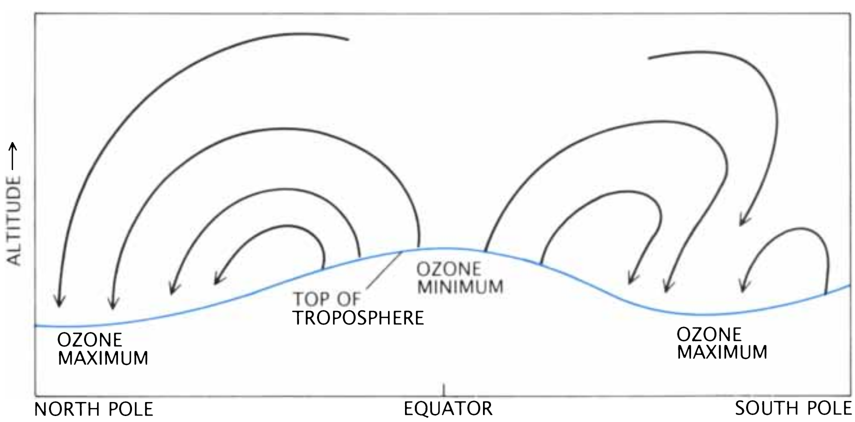
Partly as a result of such circulation patterns, the amount of ozone in the Antarctic atmosphere previously remained nearly steady at about 300 Dobson units throughout most of the winter and spring. Then the amount increased rapidly to nearly 400 Dobson units in the late spring, when the polar vortex dissipated, allowing a rapid influx of air from lower latitudes. Now the ozone amount is almost steady throughout the winter but falls rapidly in the spring to less than 200 Dobson units.

An early dynamical explanation for the springtime decline suggested that aerosols, or fine particles, from the 1982 eruption of Mexico's El Chichón volcano might have absorbed solar radiation, heating the stratosphere and causing a fountainlike movement of ozone-rich air up and out of the region. Much of the debris from the volcano has disappeared by now, however, which means ozone levels should have begun to recover. Another theory proposes that the upwelling occurs because the circulation of ozone-rich air toward the South Pole from warmer latitudes has weakened; in turn a reverse circulation has developed that moves ozone-rich polar air upward from the lower stratosphere and toward the Equator. This air would be replaced by ozone-poor air from the troposphere below.

Evidence in favor of a dynamical mechanism could come from an analysis of temperatures. If the springtime circulation has not been able to reach the polar region, one would expect to find a decrease both in heat and in ozone there. No significant changes have been discovered in the temperature of the Antarctic stratosphere in August and September, when the hole is forming, but workers have found lower average temperatures in October. The latter finding could reflect a delay in the springtime influx of warm air into the region.

As often happens in science, the meaning of the October temperature decline is open to other interpretations. It might, for instance, stem from chemically driven reductions in stratospheric ozone rather than from changes in atmospheric circulation patterns. Because ozone absorbs sunlight, the destruction of the gas in the spring could result in the absorption of less sunlight than usual and, in turn, a cooler stratosphere.

There is less equivocal evidence that some kind of dynamical process contributes to the ozone hole. The 1987 Airborne Antarctic Ozone Experiment reported that on one day, September 5, ozone levels fell by about 10 percent over an area of some three million square kilometers. The workers concluded that chemistry is not a likely explanation for such a fast and dramatic fall but that air movement is. In this instance it seems probable that ozone-poor air moved into the region temporarily, perhaps from the lower stratosphere. On the other hand, when the investigators measured the concentrations of gases that serve as tracers for air movement, they found no evidence for sustained large-scale upwelling in the stratosphere.



STRATOSPHERIC CIRCULATION, shown highly schematically, flows from the Equator downward toward the poles, carrying ozone with it. As a result global ozone levels peak not at the Equator, where the bulk of ozone is produced, but near the North Pole and at about 60 degrees south latitude. (The ozone does not travel farther south for most of the year because at that latitude the circulation encounters resistance.) The fact that ozone levels are affected by atmospheric circulation raises the possibility that the ozone hole may be caused in part by a change in airflow in the Southern Hemisphere.

Dynamical and chemical theories are also being examined as explanations for another, related scientific puzzle: the discovery that springtime amounts of stratospheric ozone have declined in the entire region south of 45 degrees in the Southern Hemisphere. A weakening of the circulation from the mid-latitudes could certainly contribute to such a decrease, but chemistry could play a role as well. For example, chemically depleted air from the polar vortex may mix with air in the surrounding area, producing a net loss of ozone.

Taken together, recent data add weight to the growing suspicion that chlorofluorocarbons do contribute importantly to the ozone hole. The findings also indicate that the phenomenon is affected by the region's unique meteorology (the polar vortex, frigid stratospheric temperatures and PSC's) and probably by a shift in airflow patterns in the Southern Hemisphere. What does all of this say about the possible danger posed to the worldwide ozone shield?

The data suggest that the extraordinary seasonal loss of ozone at the South Pole may well be a regional peculiarity that will not repeat itself in warmer climates, but this assessment is not conclusive. One thing is clear: chlorofluorocarbons are capable of altering the levels of atmospheric ozone. Moreover, the chlorine that has already been introduced into the stratosphere will interact with ozone for decades to come.

For these reasons the recent agreement to control worldwide chlorofluorocarbon consumption is laudable. Of course, debate continues about whether the treaty's targets are sufficient or unnecessarily stringent, but the issue may become clearer soon. The complete findings of the Airborne Antarctic Ozone Experiment should be available by the middle of 1988, in time for scientific review in 1989 and for a scheduled reexamination of the Montreal accord in 1990.

In the meantime there is one positive aspect to the ozone hole. In addition to convincing the international community to cooperate in reducing a threat to the environment, the hole has spurred investigators to study atmospheric chemistry and dynamics in new detail. That effort has revolutionized our knowledge of how ozone interacts with other gases and how the interactions are affected by meteorological conditions.

TOYOTA SUPRA

NERVE CENTER.
Analog instruments monitor every rev. 8-way adjustable driver's Sport Seat and optional glove-soft leather interior add even more luxury to performance.

IT DRAWS A CROWD, AND LEAVES IT BEHIND.

POWER STRUCTURE.
Good-bye crowd. Turbocharged 3.0-liter twin cam 24-valve engine with intercooler generates 230 hp at 5600 rpm.

It's an imposing presence. An eye-magnet. But the 1988 Toyota Supra Turbo is not simply for show. The low-slung aerodynamics separate it from the crowd, even at standstill. But the gap widens quickly when its electrifying turbocharged potency is released: 0 to 60 in a mere 6.7 seconds.* And control is precise: race-bred double wishbone fully independent suspension; ventilated 4-wheel power disc brakes; optional Anti-Lock Braking System (A.B.S.). Toyota Supra—#1 in its segment in customer satisfaction.** The crowd can look but they can't touch it.

Get More From Life... Buckle Up!

*United States Auto Club Certified Performance Figures.
**J.D. Power & Associates 1987 Customer Satisfaction with Product Quality and Dealer Service (for 1986 vehicles).

TOYOTA QUALITY
WHO COULD ASK FOR ANYTHING MORE!

How Killer Cells Kill

These immune-system cells recognize a target, close in on it and bind tightly to it. Then they secrete onto its surface a lethal pore-forming protein that causes the target cell to leak and die

by John Ding-E Young and Zanvil A. Cohn

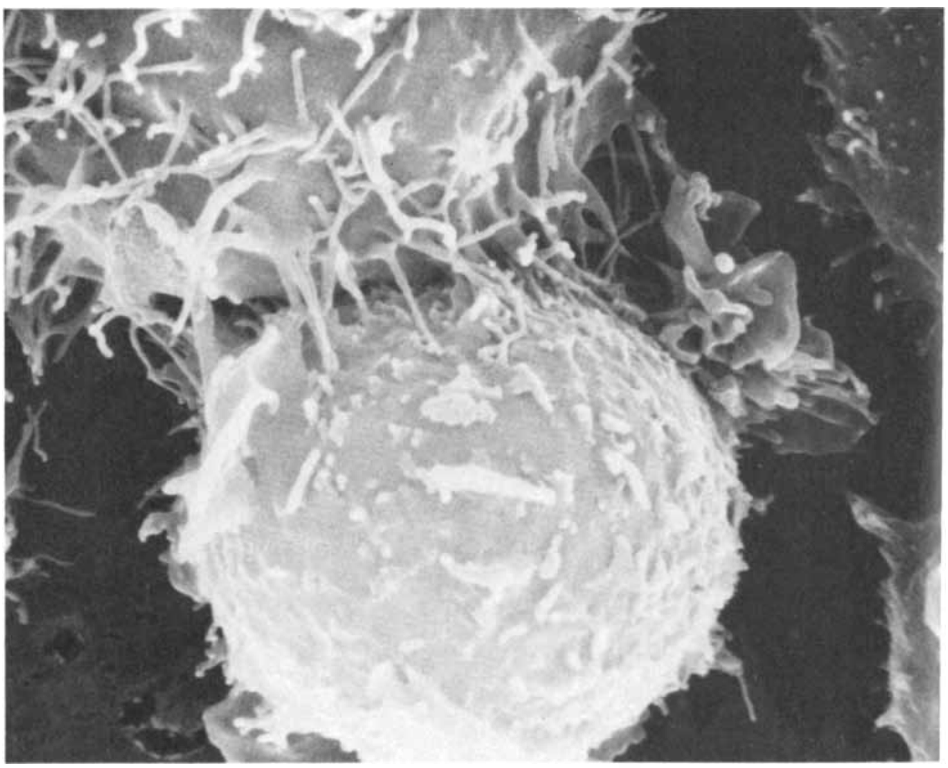
The immune system is commonly likened to an army and its various cells to soldiers. The analogy is nowhere more appropriate than in the case of the cells called killer cells. Their primary duty is to seek out and destroy the body's own cells when they go wrong: to kill tumor cells and cells that have been infected by viruses (and perhaps by other foreign agents). For some years it has been clear that killer cells do their job with great efficiency, first seeking out a miscreant target cell, then binding tightly to it and finally doing something to it that causes its death, while at the same time sparing innocent bystander cells. But what exactly do they do? Just how do the killer cells kill?

The answer is beginning to come clear. Having bound to an appropriate victim, the killer cell takes aim at the target's surface and shoots it full of holes. More specifically, it fires molecules of a lethal protein. The molecules bore into the target cell's surface membrane and form porelike channels. The target cell leaks, and soon it dies.

Work in a number of laboratories, including our own at Rockefeller University, has shown that the pore-forming protein is part of the armamentarium of the two kinds of killer cell, the cytotoxic *T* cell and the so-called natural killer cell. We have found a protein with a similar function in another immune cell, the eosinophil. Moreover, the same protein, or one very like it, appears to be responsible for attacks on human cells by an amoeba that causes severe dysentery.

We begin to think pore-forming proteins may be a major weapon in a wide range of cell-mediated killing. Knowing more about the process should have important medical payoffs. It may eventually be possible, for example, to treat amebic dysentery and some other parasitic, fungal and bacterial diseases by blocking a pore-forming protein. Finding a way to enhance the pore-forming process in immune-system cells might be even more important: in principle it should be helpful in the treatment of both cancer and such intractable viral diseases as AIDS.

The killer cells are elements of the cellular immune system, but their precise and lethal functioning can be studied and understood only in the context of the immune system as a whole. It has a humoral component as well as a cellular one. The humoral system defends the body primarily against bacteria and toxic molecules. Its weapons are antibodies, or immunoglobulins, which are synthesized and secreted by the cells called *B* lymphocytes. Each of millions of *B* cells synthesizes a particular antibody, which recognizes a particular antigen (a molecular recognition pattern); the cell displays the antibody on its surface. When the lymphocyte encounters a bacterial



KILLER CELL is seen destroying a tumor cell in this series of scanning electron micrographs. In the micrograph at the left a cytotoxic *T* cell (*top*) makes contact with the smaller target cell. In the middle micrograph the damage has been done. The target cell, its outer membrane shot full of holes by a protein secreted from the killer cell, has be-

cell or a toxin bearing that antigen, the lymphocyte proliferates. Some of its progeny become memory cells that will respond to the same antigen faster the next time; most of the progeny become plasma cells that manufacture a large amount of the antibody and secrete it. The antibody binds to the antigen. Toxins are precipitated or otherwise neutralized by the binding itself. In the case of an invading cell the binding sets off a cascade of reactions at the cell surface involving a group of blood-serum proteins collectively known as complement. The end result of the cascade is the death of the cell.

The same precursor cells that give rise to *B* lymphocytes are the ancestors of a varied family of *T* lymphocytes that are the basis of the cellular immune system. Some *T* cells, called helper cells and suppressor cells, modulate both the humoral and the cellular system, chiefly by secreting chemical messengers called lymphokines. The major effector cell of the cellular system is the cytotoxic *T* lymphocyte, or killer *T* cell, we have mentioned. Its main targets are virus-infected cells. The other type of killer cell introduced above, the natural killer cell, is also a lymphocyte. Its precise lineage is not clear, but it seems to be closely related to the cytotoxic *T* cell. Its main targets are thought to be tumor cells, and perhaps also cells infected by agents other than viruses.

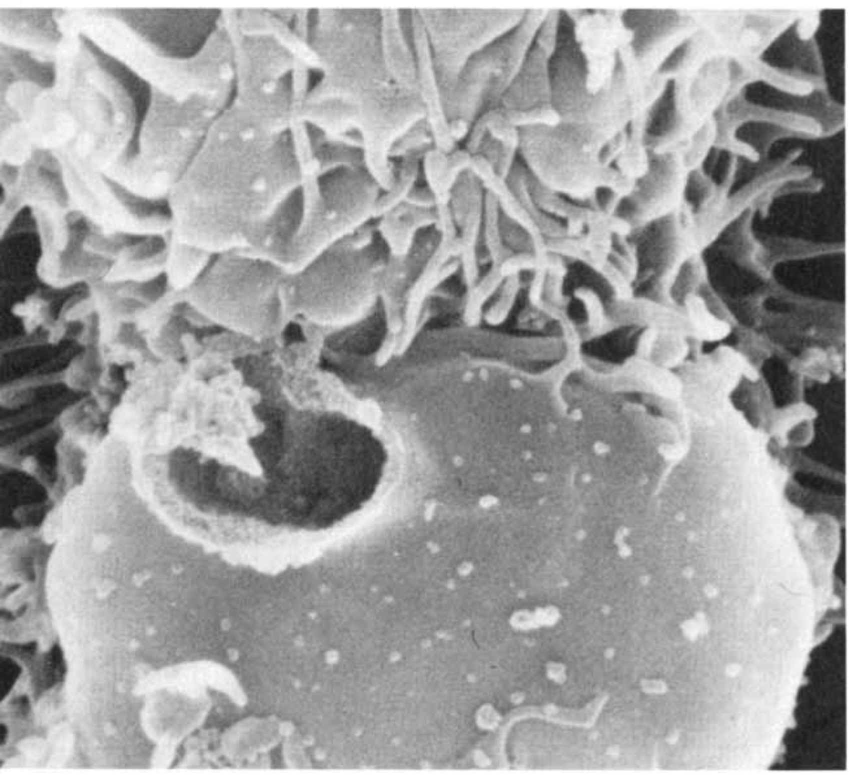
As in the case of *B* lymphocytes, the function of *T* lymphocytes depends in the first instance on proper recognition of an appropriate target. *T* cells display on their surface specific receptors, very similar to the *B* cell's antibodies, that recognize and bind to particular cell-surface antigens. *T* cells are more selective than *B* cells, however. They recognize and bind to an antigen only when it is "presented" by one of a group of surface molecules known collectively as the major histocompatibility complex (MHC). Natural killer cells, as their name implies, are less choosy about what they attack. Their receptors are less discriminating, and they are not restricted by the MHC.

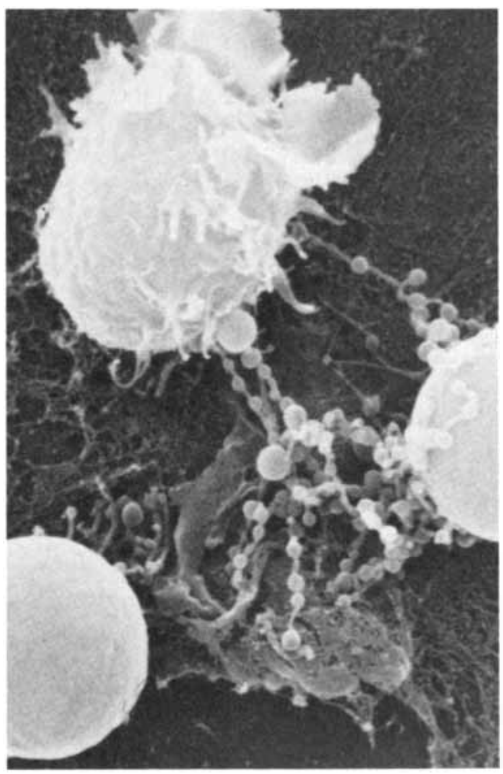
Once either the killer *T* cell or the natural killer cell has identified its target, the killer cell binds tightly to the target cell. This close contact triggers the lethal process and also ensures that neighboring cells are not indiscriminately destroyed.

That much was known a decade ago, but the nature of the killing process continued to be a mystery. The first clues to the mystery came in the early 1970's from the laboratories of a number of investigators, notably Eric Martz of the University of Massachusetts at Amherst, Christopher S. Henney of the Immunex Corporation in Seattle, William R. Clark of the University of California at Los Angeles, Pierre Golstein of the Center for Immunology in Marseilles and Gideon Berke of the Weizmann Institute of Science in Israel. Their work dissected the killing process into a sequence of discrete stages.

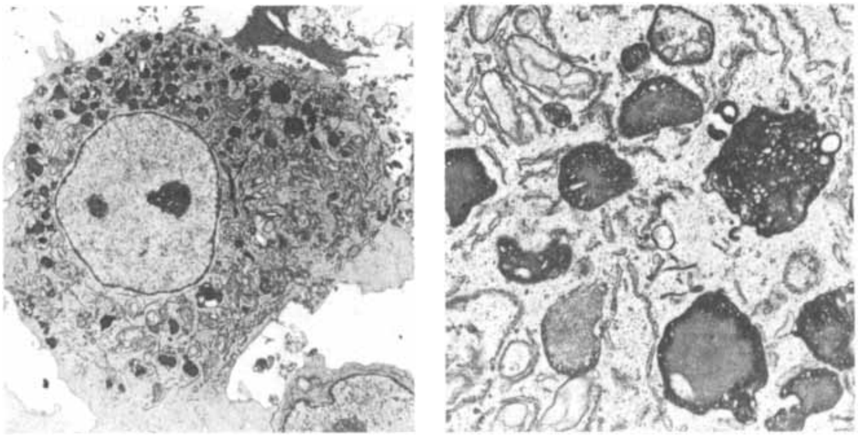
First, it became clear, a "conjugate" is formed as the lymphocyte and its target come in close contact. Then some kind of lethal hit is delivered, injuring the target cell. The hit initiates what seems to be a programmed death of the injured target: death proceeds in a predetermined manner (provided only that calcium ions are present) even as the conjugate breaks up and the lymphocyte goes on to initiate a new killing cycle.

Henney and Martz were among the first to suggest that a lymphocyte appeared to kill by somehow damaging the plasma (outer) membrane of





come leaky and an influx of water has made it balloon; it has lost many of its projecting villi and there is a large cavity in the membrane. In the micrograph at the right there is nothing left of the tumor cell but its nucleus and some debris (the other nucleus in the micrograph is that of a second target cell). The cells were prepared by Chau-Ching Liu and the scanning microscopy was done by Gilla Kaplan. The cells are enlarged 5,700 diameters in the first two micrographs and 4,200 diameters in the third one.



CULTURED CYTOTOXIC *T* CELL is enlarged 1,400 diameters in an electron micrograph (*left*) made by the authors with Hans Hengartner and Eckhard R. Podack. The dark storage granules in its cytoplasm, enlarged 7,600 diameters (*right*), have an electron-dense outer membrane; their amorphous matrix contains perforin, the lethal protein.

its target. Their proposal was based on their observation that radioactive molecules introduced into a target cell as markers leaked out rapidly when the target was damaged by a lymphocyte. The membrane became permeable only to markers of a certain maximum size, suggesting that the damage to the membrane might take the form of holes, or pores.

That possibility became a likelihood in 1980. Robert R. Dourmashkin and Pierre Henkart of the National Cancer Institute and their colleagues examined the surface of damaged target cells, enlarged many thousands of diameters in electron micrographs, and detected ringlike structures that appeared to be holes in the membrane. Their finding was extended three years later by Eckhard R. Podack of the New York Medical College and Gunther Dennert of the University of Southern California School of Medicine, who studied the effect of cultured killer cells on tumor cells. They established that the surface of the target cell was pocked with holes whose internal diameter ranged from five to 20 nanometers (millionths of a millimeter).

It was still not at all clear whether the pores were actually inflicted by the lymphocytes or simply reflected a terminal stage of cell death caused by some other form of injury. The search for an answer was long hampered by the lack of a ready source of killer cells. Then in 1977 Steven Gillis of Immunex, Kendall A. Smith of the Dartmouth Medical School and Robert C. Gallo's group at the National Cancer Institute succeeded in maintaining mouse lymphocytes—both cytotoxic *T* cells and natural killer cells—in the laboratory. They were able to do so by identifying particular nutrients and growth factors required for the survival of these cells in culture; one key factor turned out to be the lymphokine called interleukin-2. Culturing made it possible to grow lymphocyte clones derived from known parent cells, and so to have at hand an abundant and homogeneous source of cytotoxic *T* cells and natural killer cells with known characteristics. The way was cleared for detailed analysis of such cells with the tools of cell biology and biochemistry.

An obvious feature of the lymphocytes clearly called for investigation. Micrographs had revealed numerous small, dark organelles (subcellular elements) in the cytoplasm. They appeared to be storage granules, which are common in secretory cells. Such granules provide an efficient means of accumulating and packaging a substance synthesized by the cell so that large quantities of it can be released quickly at the right time. The release is accomplished by exocytosis: the granules move to the periphery of the cell, fuse with the plasma membrane and disgorge their contents.

Several investigators had seen that early in the cell-killing process the granules in a killer lymphocyte become concentrated in the part of the cell that is in close contact with the target. Then Abraham Kupfer and S. Jonathan Singer of the University of California at San Diego and Dennert noted that the granule-packaging Golgi apparatus seems to be directed toward that contact region and that a number of proteins of the cell's cytoskeleton (its fibrous internal framework) "reorient" toward the target cell soon after contact is established, apparently providing the motive force that redirects the Golgi and the granules.

The reorientation of the cytoskeleton and the movement of the Golgi stacks and granules take place only when and where the lymphocyte binds to an appropriate target. Roger Y. Tsien of the University of California at Berkeley showed that the binding induces an explosive increase of calcium ions in the cell; the increase triggers exocytosis. John H. Yanelli and Victor H. Engelhard of the University of Virginia have made cinemicrographs showing granules reorienting in the cytoplasm and then fusing with the plasma membrane.

All in all, the evidence suggested that contact with an appropriate target causes the killer cell to aim its secretory apparatus at the target and fire a lethal agent contained in its granules. It was necessary first to establish that the granules are in fact the shell casings and then to identify the projectile itself.

The first task was to isolate the granules and see if they alone could kill. That was done independently in 1984 by Henkart and Podack and by our group. We exploited the various techniques of subcellular fractionation, the objective of which is to break up a cell into its components and find what component contains a particular enzyme or carries out a particular function. Killer lymphocytes are broken into bits and pieces by subjecting them to pressure in nitrogen gas. The cellular debris is layered onto a density gradient of inert particles and then spun in a high-speed centrifuge. The various organelles come to rest, depending on their density, in discrete bands. We examined each of the fractions in the electron microscope and assayed them for enzymatic activity and ability to kill cells.

One fraction, which electron micrography showed consisted almost entirely of granules, was enriched in certain enzyme activities and was a potent killer: when the isolated granules were mixed with red blood cells or tumor cells in a medium containing calcium ions, the cells died within minutes. Electron micrographs revealed that the cells' surface carried ringlike lesions virtually indistinguishable from those produced by intact killer cells. And so the granules were shown indeed to contain the killer cells' lethal secretion.

The secreted agent itself was soon identified. In 1985, collaborating with Podack and Hans Hengartner of University Hospital in Zurich, we found a protein that all by itself (in the presence of calcium) reproduces the observed membrane lesions and the killing inflicted by intact killer cells and by granules. We isolated and purified the protein by passing extracts of granules through chromatographic columns that separated proteins on the basis of electric charge and of molecular mass, and screening the separated proteins for their efficiency in lysing red blood cells: breaking down their membrane and causing them to burst. The killer protein was isolated independently by Danielle Masson and Jürg Tschopp of the University of Lausanne.

So far only this one pore-forming protein, which is often called perforin (because it perforates membranes), has been found in the granules of cytotoxic *T* lymphocytes or natural killer cells. Its molecular mass is 70,000 daltons. Once cells are exposed to perforin in the presence of calcium ions they are lysed within a few minutes. On the other hand, if calcium ions are added to perforin before it makes contact with cells, the protein's killing activity is abolished. The effect sounds paradoxical, but actually it leads to an important insight into just how perforin kills.

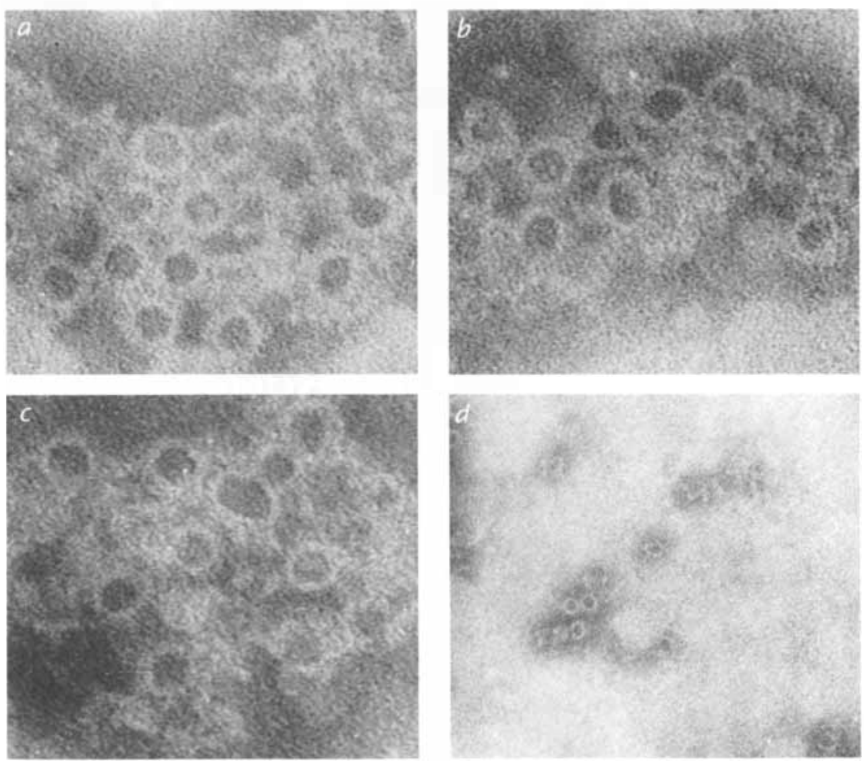
The 70-kilodalton molecules secreted by a killer cell insert into the target-cell membrane. There (in the presence of calcium ions) the individual molecules (monomers) polymerize, or link up with one another. The polymer they form can assume various shapes, but under optimal conditions the final product resembles a cylinder. In the electron microscope it looks like a ring when seen in cross section and like two parallel lines when seen in longitudinal section. A fully formed ring, as Podack and Dennert first noted, ranges from five to 20 nanometers in internal diameter.

For perforin to damage a target cell the calcium-mediated polymerization must take place entirely within the cell's membrane. The reason is that only the perforin monomer can insert into a membrane; if polymerization takes place in solution, in the absence of nearby membrane, the resultant polymer cannot enter the membrane and cannot kill. The protective effect is easy to understand. Any secreted perforin that spills into the extracellular space or the bloodstream, where calcium is abundant, must immediately undergo polymerization and thereby be rendered inactive, virtually eliminating the possibility of "accidental" injury to nontarget cells.

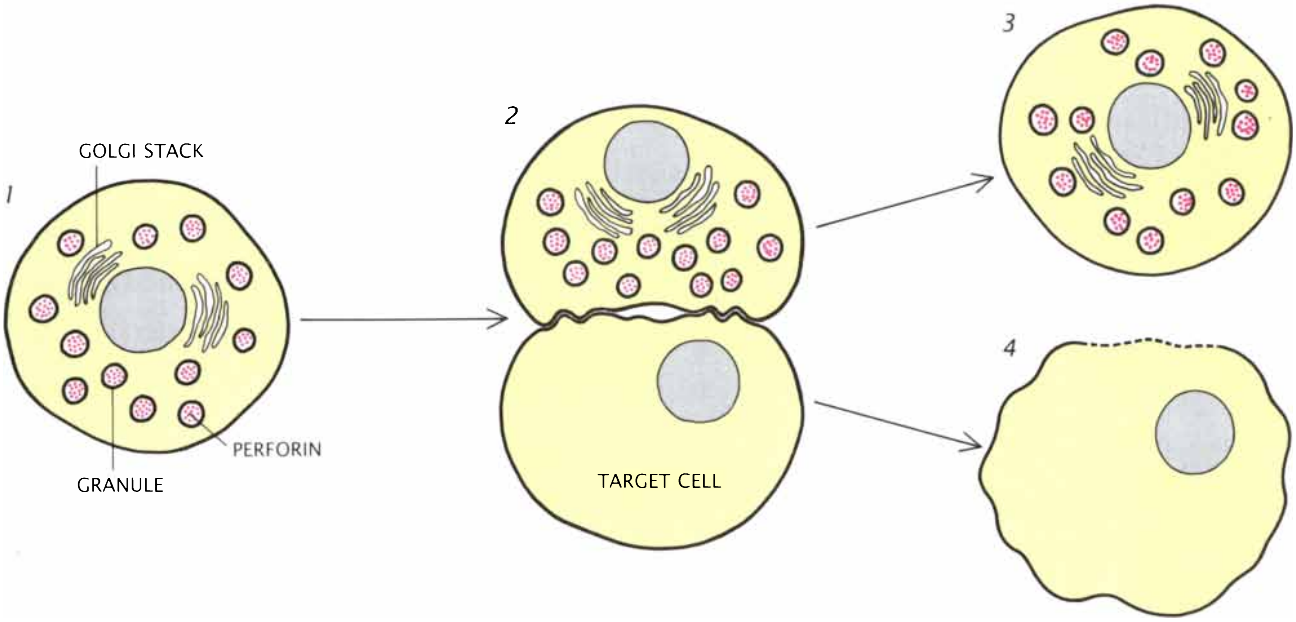
On an appropriate target cell, however, the tubular pores shaped by polymerization bring about rapid, measurable changes in the target cell. The plasma membrane of a living cell retains proteins and other large molecules within the cell (except when they are to be secreted) and segregates different ions, keeping some inside and some outside the cell. The segregation of positive and negative ions establishes a transmembrane electric potential. When the membrane leaks, ions and water tend to flow down their electrochemical gradients toward equilibration, and so there is a drop in membrane potential. If the holes in the membrane are of limited size, there is an additional effect, known as the Donnan effect. The large molecules inside the membrane cannot pass through it; water and salts from the extracellular fluid pour through the membrane toward the side that has the proteins—into the cell. The cell swells, and eventually it bursts.

By impaling target cells with microelectrodes, we were able to measure a significant drop in membrane potential soon after perforin was administered. With sensitive electronic equipment we were even able to measure the ion current flowing through individual pores. Our results were consistent with the predictions of the Donnan effect. The measurements showed, moreover, that the perforin holes are stable structures that persist as open pathways in the membrane. By determining just which ions and small molecules got through the damaged membrane, we were able to estimate that the functional internal diameter of the pores ranges from one to 10 nanometers (as opposed to the observed diameter in micrographs of from five to 20 nanometers).

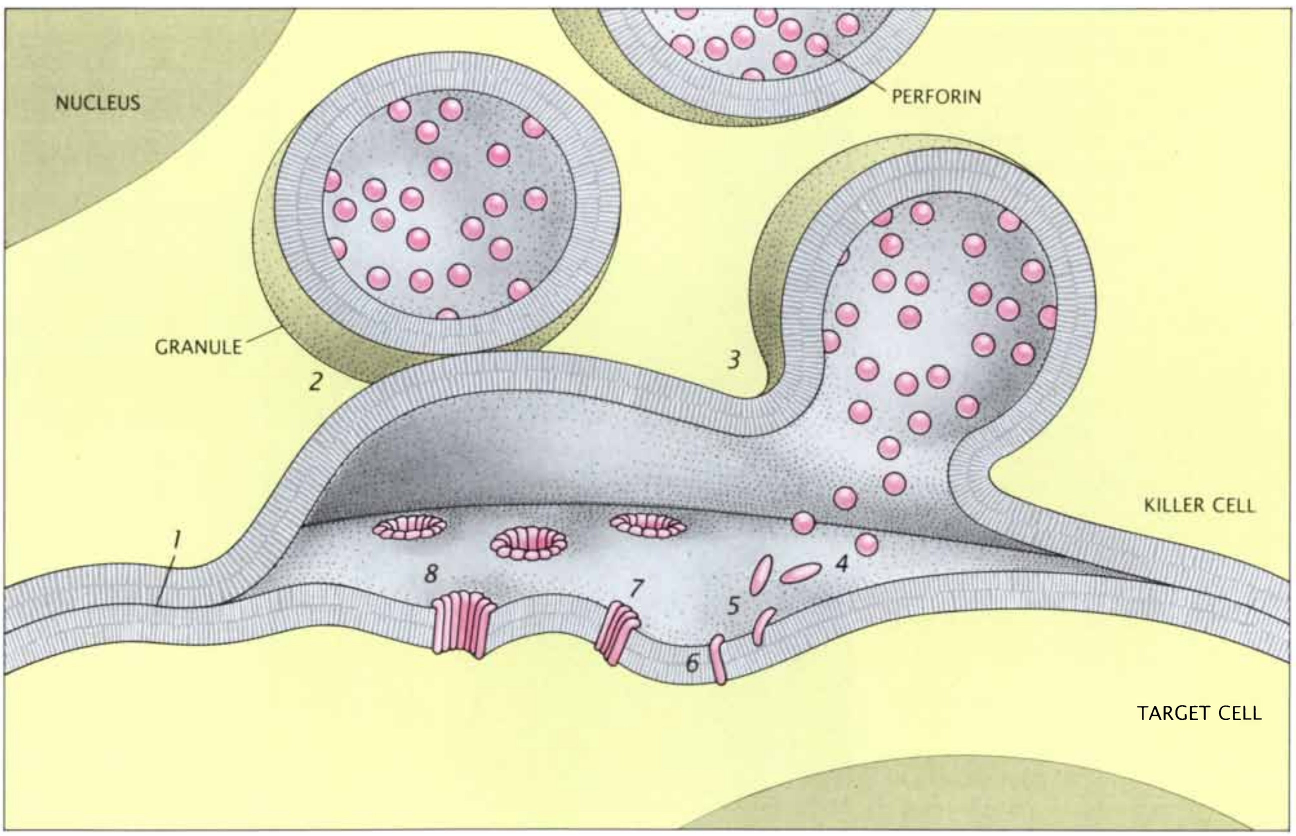
Given the apparently undiscriminating efficiency of perforin as a killer, how can one account for the specificity of killing by lymphocytes? We have mentioned that "accidental" killing by perforin that gets



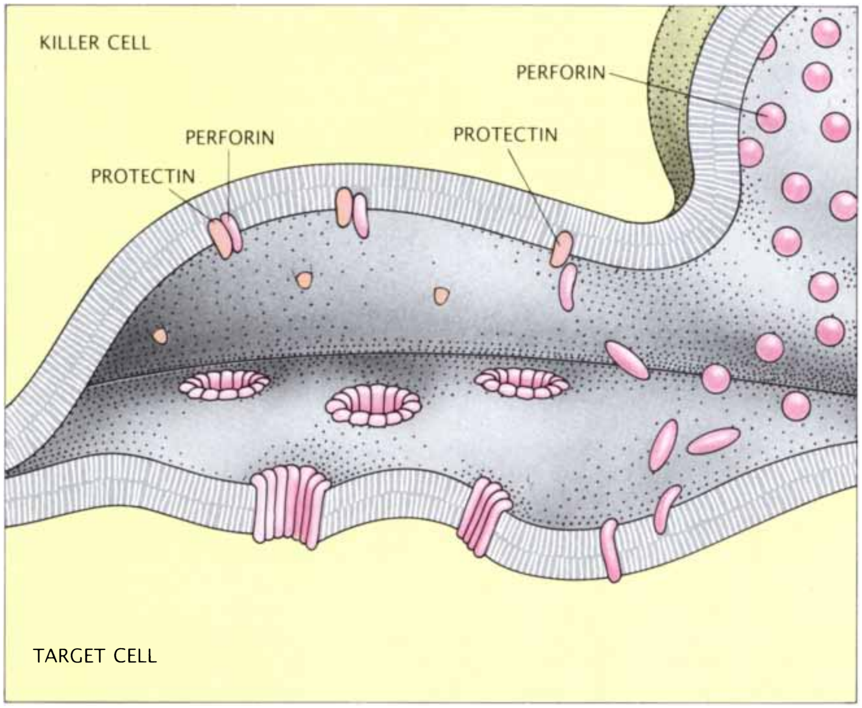
FATAL LESIONS are tubelike pores with an internal diameter of from five to 20 nanometers (millionths of a millimeter), visible in electron micrographs of negatively stained membrane as rings or parallel lines depending on orientation. The lesions look about the same whether inflicted by intact lymphocytes (*a*), by isolated granules (*b*) or by perforin purified from the granules (*c*). Smaller pores (two to three nanometers in diameter) are made by a protein isolated from the amoeba *Entamoeba histolytica* (*d*).



STAGES IN KILLING PROCESS were identified some time ago. The killer lymphocyte (*1*) recognizes the target cell and makes close contact with it (*2*). On contact the cell's granules and the Golgi complex that forms them are reoriented toward the target cell; perforin is secreted and forms pores in the target-cell membrane (*see illustration below*). Having launched its lethal missiles, the killer cell withdraws and goes on to kill again (*3*); the damaged target cell dies a programmed death within minutes (*4*).



DETAILS OF KILLING PROCESS have now been elucidated. A rise in the lymphocyte's calcium-ion level, apparently triggered by the receptor-mediated binding of the killer cell to its target (*1*), brings about exocytosis, in which the granules fuse with the cell membrane (*2*) and disgorge their perforin (*3*) into the small intercellular space abutting the target. Calcium there changes the conformation of the individual perforin molecules, or monomers (*4*), which then bind to the target-cell membrane (*5*) and insert into it (*6*). The monomers polymerize like staves of a barrel (*7*) to form pores (*8*) that admit water and salts and kill the cell.



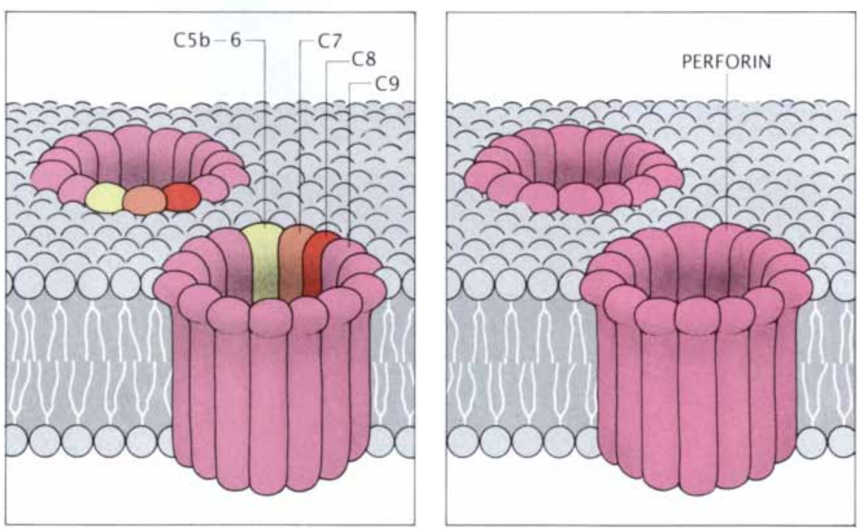
PROTECTIVE MECHANISM that prevents self-inflicted death of the killer cell has been postulated by the authors. They think a protein, protectin, that is very similar to perforin is concentrated in the lymphocyte membrane; perforin monomers bind to the protectin and are prevented from polymerizing to form pores in the killer-cell membrane.

away from the lymphocyte-target interface is prevented by calcium-induced polymerization of the protein. What one might term "purposeful" killing of nontarget cells (as a result of contact with a killer lymphocyte) is prevented, on the other hand, by the lymphocyte's ability to recognize an appropriate cell—one displaying, for example, either viral antigens or tumor antigens. In other words, the specificity of killing resides solely in the requirement for close contact, which depends in turn on the binding of killer cell to target cell through the interaction of antigens on the target cell with receptors on the killer cell.

But what keeps the killer cell from killing itself? It cannot be the close-contact requirement, since the killer cell's membrane is continually exposed directly to secreted perforin. Chau-Ching Liu, a graduate student in our laboratory, and one of us (Young) collaborated with Clark recently to show that even purified perforin fails to kill either cytotoxic *T* cells or natural killer cells. The self-protective mechanism is not known, but we have a hypothesis. We think the killer-cell membrane may incorporate a special protein, which we call protectin, that is very similar to perforin. The close homology would promote a kind of "faulty polymerization": the protectin would rapidly combine with any perforin monomer that gets to the killer-cell membrane, thereby preventing either the insertion of perforin into the membrane or the normal perforin-to-perforin polymerization that would form a pore. We are now engaged in an intensive search for our hypothetical protein.

All the recent studies we have described here were done with mouse-lymphocyte cultures maintained in the laboratory. It was conceivable that perforin was just a laboratory curiosity and not truly the lymphocytes' weapon in vivo. Collaborating with Bice Perussia of the Wistar Institute in Philadelphia and Liu, we looked for perforin in lymphocytes freshly obtained from human blood. We could find none. When the lymphocytes were stimulated with interleukin-2, however, they proliferated in vitro and began to synthesize perforin. We found this to be the case for both cytotoxic *T* cells and natural killer cells; similar results have been reported by Leora S. Zalman and Hans J. Muller-Eberhard of the Research Institute of Scripps Clinic and their colleagues. The in vitro effect of interleukin-2 presumably reflects its effect in the body, where it is produced by helper *T* cells and promotes a range of immune responses.

Indeed, the effect we noted in the laboratory may explain an apparent clinical effect of interleukin-2 first reported in 1984 by Steven A. Rosenberg of the National Cancer Institute. He devised a novel therapy for certain intractable cancers in which lymphocytes extracted from the blood of a patient are stimulated with interleukin-2 outside the body and then infused back into the patient.



HUMORAL IMMUNE SYSTEM forms pores much like those inflicted by the cellular system's killer lymphocytes. Binding of antibodies to a target cell triggers a cascade in which successive proteins of the complement system are activated. Eventually the protein C5b-6 binds to the target's surface membrane, after which C7, C8 and a number of C9 proteins aggregate to form a pore (*left*). In contrast, the pores made by killer cells are formed by the self-aggregation of one kind of subunit: the perforin monomer (*right*).

The lymphocytes are presumably thereby activated to kill more efficiently; some tumor regression has been observed in some patients. The procedure is highly toxic and therefore still in an experimental stage, but a better understanding of how to induce the expression of perforin by killer lymphocytes will certainly have a role in designing immunotherapies for cancer.

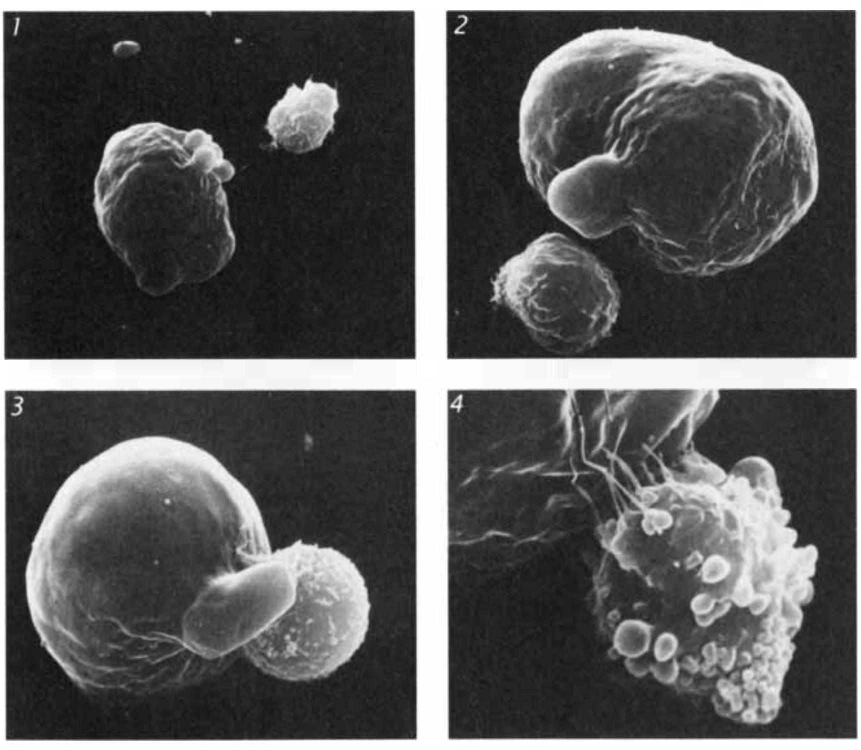
Although it seems clear that killer cells kill by punching holes in target-cell membranes, the lymphocytes may have other weapons too. John H. Russell of Washington University in St. Louis has put forward an "internal disintegration" model to explain the killing event. The model is based on the observation that early in the course of target-cell injury the membrane enclosing the nucleus of the cell ruptures and the DNA in the nucleus breaks up into small fragments. Russell and others argue that the death of the target cell results from the cleaving of the DNA, which is presumably triggered by some signal emitted by the killer cell. The model has not been elaborated in detail, but it cannot be excluded by our findings. Perhaps there are several different killing mechanisms.

Indeed, we have found that certain cytotoxic *T*-cell lines maintained in culture for some time continue to exert killing activity even though they do not secrete perforin. Do they secrete something else? We have preliminary evidence that a molecule we call leukalexin, which kills cells over the course of several hours rather than minutes, may be present in killer cells. It is just possible that even if perforin pores first admit only water and salts and are too small to pass most proteins, they eventually enlarge enough (through additional polymerization) to admit one chain of a protein—perhaps leukalexin—that cleaves DNA. Alternatively, pore formation might lead eventually to the admission of a DNA-cleaving protein by somehow promoting endocytosis, the process by which large molecules are generally taken into cells.

Pore formation may not be the only killing mechanism for killer lymphocytes, but it is clearly an efficient one. It figures, for example, in humoral immunity as well as in cellular immunity: it is the end result of the complement cascade that kills bacteria labeled by antibodies. The terminal proteins of the cascade polymerize to form holes, much like the pores made of perforin, with an internal diameter of 10 nanometers. Podack and we have found, as has Tschopp independently, that the terminal complement proteins have significant homology with perforin: the sequence of amino acids (the building blocks of proteins) is identical in some parts of these complement proteins and of perforin. This similarity between elements of the humoral and the cellular immune systems cannot have come about by accident. We speculate that the killer proteins of both systems had a common ancestry but diverged somewhat to become specialized for their respective roles. Once an efficient mechanism evolves in an organism, it tends to be well conserved by natural selection.

Cell killing is not limited to immune cells, or indeed to organisms of higher complexity. Numerous species, including certain bacteria, fungi and protozoan parasites, are effective killers of cells. Among the weapons many of them have in common are proteins that punch holes in the surface of a target cell. We chose to study one such species in an attempt to improve our understanding of how lymphocytes kill their targets. Certain virulent strains of the amoeba *Entamoeba histolytica* infect human beings worldwide, invading the intestine to cause severe dysentery and often spreading to other organs. In a laboratory culture these strains kill a wide range of cells, in each case first binding tightly to the target.

In 1982 Carlos Gitler of the Weizmann Institute and we independently isolated from *E. histolytica* a protein that forms pores in the surface membrane of target cells. The purified pore-forming protein is a potent killer. This PFP, as we first named it, is much smaller than perforin (14,000 daltons), but like perforin it polymerizes and forms large tubular lesions, with an internal diameter of from two to three nanometers, in a cell membrane. We think the amoeba kills by binding to its target and shooting it full of holes made of a PFP polymer. The similarity of the killing mechanisms in killer cells that attack the body and killer cells that defend it would seem to be a nice example of convergence in evolution. The amoeba and the human lymphocytes have independently developed functionally similar pore-forming proteins that accomplish the same objective: the death of cells.



PARASITIC AMOEBA can kill with a pore-forming protein that has the same effect as perforin. In these scanning electron micrographs made by Gilla Kaplan *Entamoeba histolytica* (the larger cell) approaches its target, an immune-system macrophage (*1*), and extends a pseudopod (*2*). After making contact the amoeba may kill its target by phagocytosis, or ingestion (*3*), or instead may secrete its lethal pore-forming protein, which causes the target to swell and show extensive blebbing, or blistering (*4*), and then to die.

A BLACK HOLE THAT HELPS MILITARY HELICOPTERS ESCAPE HOT PURSUIT.

Because of hot engine exhaust, military helicopters can be easy prey for heat-seeking missiles. The best way to escape this threat and still accomplish the mission: Cool the exhaust. To do this, our engineers developed the Black Hole Infrared Suppression System, currently in use on the U.S. Army Apache.

Engine exhaust is mixed with outside air, routed through a series of pipes and expelled through finned nozzles. The exhaust is so cool it is ignored by heat-seeking missiles.

Because it stays cool, the Black Hole exhaust system can be made of inexpensive metals. And because the system has no moving parts, it requires little maintenance.

We're giving America its money's worth with breakthroughs in aviation, space and information systems.

We're McDonnell Douglas.

For more information, write:
Apache, McDonnell Douglas,
Box 14526, St. Louis, MO 63178.

MCDONNELL DOUGLAS

HELICOPTERS SPACECRAFT & MISSILES MILITARY & COMMERCIAL AIRCRAFT INFORMATION SYSTEMS TRAINING SYSTEMS HEALTH CARE FINANCING

The Reality of the Quantum World

Einstein held that quantum-mechanical descriptions of physical systems are incomplete. Laboratory tests show he was probably wrong; the bizarre nature of the quantum world must be accepted

by Abner Shimony

We live in a remarkable era in which experimental results are beginning to elucidate philosophical questions. In no domain have the results been more dramatic than in quantum mechanics. The theory has been confirmed magnificently since the 1920's as its predictions of atomic, molecular, nuclear, optical, solid-state and elementary-particle phenomena were shown to be accurate. Yet in spite of these successes the bizarre and counterintuitive character of quantum mechanics has led some investigators, including Einstein, to believe quantum-mechanical descriptions of physical systems are incomplete and in need of supplementation. Recent experiments show that this opinion is very likely wrong. The experimental results reveal more clearly than ever that we live in a strange "quantum world" that defies comfortable, commonsense interpretation.

Here are a few of the new, strange findings we must begin to accept. First, two entities separated by many meters and possessing no mechanism for communicating with each other nonetheless can be "entangled": they can exhibit striking correlations in their behavior, so that a measurement done on one of the entities seems instantaneously to affect the result of a measurement on the other. The finding cannot be explained from a classical point of view, but it agrees completely with quantum mechanics. Second, a photon, the fundamental unit of light, can behave like either a particle or a wave, and it can exist in an ambiguous state until a measurement is made. If a particlelike property is measured, the photon behaves like a particle, and if a wavelike property is measured, the photon behaves like a wave. Whether the photon is wave- or particlelike is indefinite until the experimental arrangement is specified. Finally, the notion of indefiniteness is no longer confined to the atomic and subatomic domains. Investigators have found that a macroscopic system can under some circumstances exist in a state in which a macroscopic observable has an indefinite value. Each of these findings alters drastically the way we perceive the world.

An understanding of these experiments and their philosophical implications requires some familiarity with the basic ideas of quantum mechanics. Essential to any discussion of the theory is the concept of the quantum state, or wave function. The quantum state specifies all the quantities of a physical system to the extent that it is possible to do so. The caveat at the end of the preceding sentence is crucial, because according to quantum mechanics not all quantities of a system have simultaneously definite values. The familiar Heisenberg uncertainty principle, which asserts that the position and the momentum of a particle cannot be simultaneously definite, is perhaps the best-known instance of this proposition.

What the quantum state of a system does provide unequivocally is the probability of each possible outcome of every experiment that can be done on the system. If the probability is 1, the outcome is certain to occur; if the probability is zero, the outcome is certain not to occur. If, however, the probability is a number between zero and 1, then it cannot be said in any individual case what the outcome will be. All that can be said is what, on the average, the outcomes of a specified experiment carried out on a large number of replica systems will be.

Imagine, for instance, that measurements are made on a photon. The quantum state of the photon is fixed if three quantities are known: the photon's direction, its frequency and its linear polarization (the direction of the electric field associated with the photon). A suitable apparatus for measuring polarization is a sheet of polarizing film. The film is idealized so that it transmits all light incident on it at a right angle if the light is linearly polarized along a certain direction in the film called the transmission axis. The film blocks all light incident on it at a right angle if the light is linearly polarized perpendicular to the transmission axis.

Various experiments can be performed by rotating the polarizing film in different ways. If the photon is linearly polarized along the transmission axis, there is a probability of 1 that it will be transmitted. If the photon is linearly polarized perpendicular to the transmission axis, the probability that it will be transmitted is zero. A further implication of quantum mechanics, going beyond what has been said so far, is that if the photon is linearly polarized at some angle to the transmission axis between zero and 90 degrees, the probability of transmission is a number between zero and 1 (specifically, the square of the cosine of that particular angle). If the probability is, say, one-half, then out of 100 photons linearly polarized at the corresponding angle to the transmission axis 50 will be transmitted on the average.

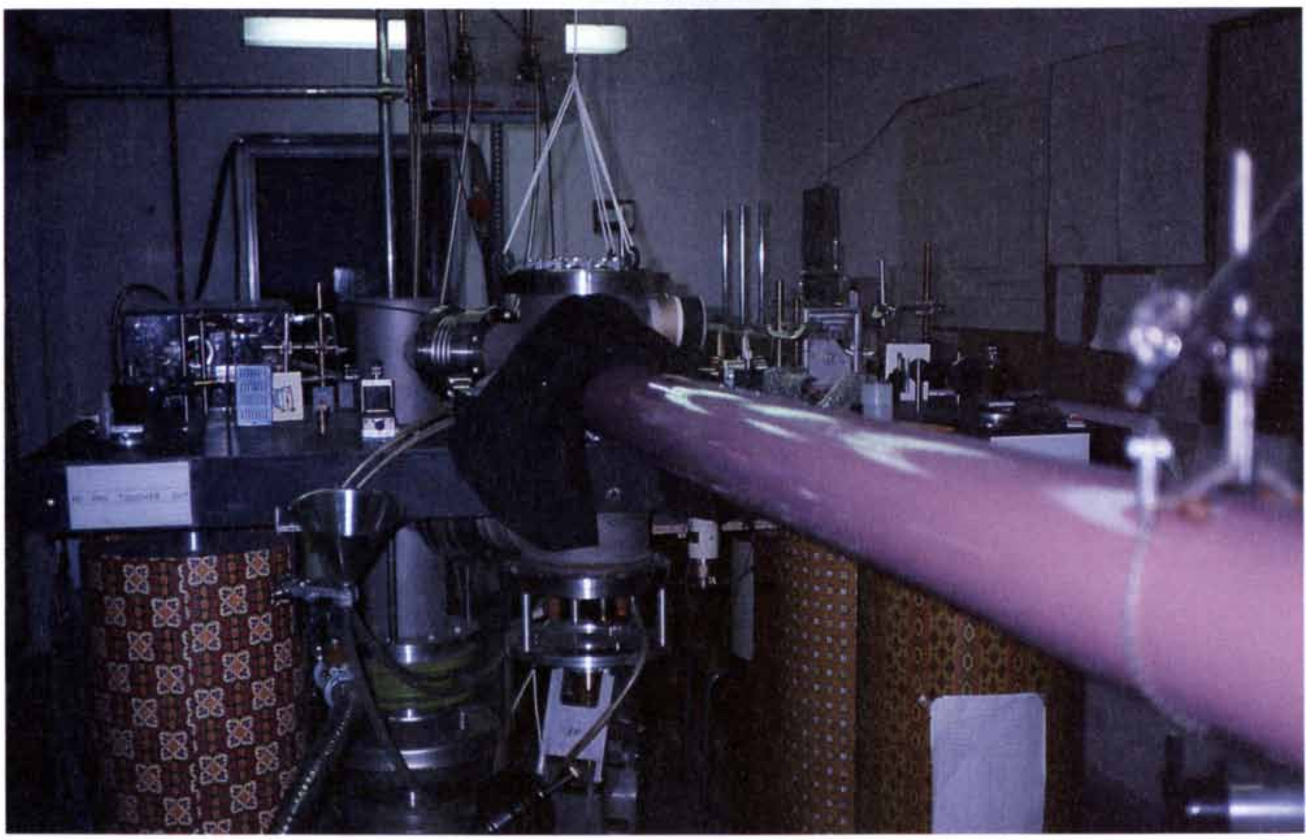
Another basic idea of quantum mechanics is the superposition principle, which asserts that from any two quantum states of a system further states can be formed by superposing them. Physically the operation corresponds to forming a new state that "overlaps" each of the states from which it was formed. The concept can be illustrated by considering two quantum states of a photon in which the direction of the photon's polarization in the first state is perpendicular to the direction of the photon's polarization in the second. Then any number of states can be formed in which the photon's polarization points at some angle between the two perpendicular directions.

From these two basic ideas alone—indefiniteness and the superposition principle—it should be clear already that quantum mechanics conflicts sharply with common sense. If the quantum state of a system is a complete description of the system, then a quantity that has an indefinite value in that quantum state is objectively indefinite; its value is not merely unknown by the scientist who seeks to describe the system. Furthermore, since the outcome of a measurement of an objectively indefinite quantity is not determined by the quantum state, and yet the quantum state is the complete bearer of information about the system, the outcome is strictly a matter of objective chance—not just a matter of chance in the sense of unpredictability by the scientist. Finally, the probability of each possible outcome of the measurement is an objective probability. Classical physics did not conflict with common sense in these fundamental ways.

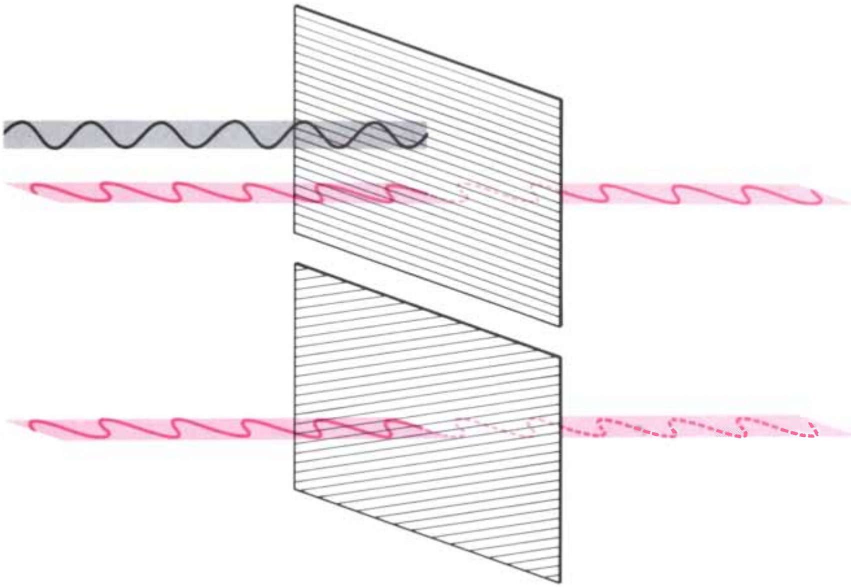
Even more startling implications flow from quantum mechanics if the system consists of two correlated parts. Suppose two photons fly apart in opposite directions. One possible quantum state of the pair of photons is the state in which both photons are linearly polarized along a vertical axis. Another possible state is the one in which they are both linearly polarized along a horizontal axis. There is nothing particularly bizarre or surprising about either of these two-photon quantum states, beyond the peculiarities of the single-photon states mentioned above. But if the superposition principle is brought into play, strange effects can occur.

In particular, by using the superposition principle one can form a quantum state that contains equal amounts of the vertically polarized state and the horizontally polarized state. This new state will figure prominently in what follows, and so it will be given a name, Ψ_0 (since the Greek letter psi is commonly used to represent a quantum state). The properties of Ψ_0 are most peculiar indeed. Imagine, for instance, inserting in the paths of the photons polarizing films with vertically oriented transmission axes. Because Ψ_0 contains equal amounts of the vertically and horizontally polarized states, there is a probability of one-half that both photons will be transmitted through their respective films and a probability of one-half that both will be blocked. What cannot happen is that one photon will be transmitted and the other will be blocked. In other words, the outcomes of the linear-polarization experiments on the two photons are strictly correlated.

The results will be the same if the polarizing films are oriented at an angle of 45 degrees with respect to



EXPERIMENTAL TESTS are now shedding light on topics in quantum mechanics that were once confined to the realm of philosophical debate. In this experiment, which was done by Alain Aspect and his colleagues at the Institute of Optics of the University of Paris, the lasers at each side of the picture excite individual calcium atoms in the vacuum chamber (*center*). Each atom returns to its unexcited state by emitting a pair of photons. (The photon is the fundamental unit of light.) The photons travel in opposite directions through 6.5 meters of pipe, and those that pass through polarization analyzers impinge on photodetectors. Quantum mechanics predicts there should be delicate correlations in the polarizations of the oppositely directed photons; the correlation conflicts with classical theories called hidden-variables models. The experiment confirmed quantum mechanics.



INDEFINITENESS of a quantum system is illustrated for a photon. A sheet of polarizing film transmits all light incident on it at a right angle if the light is linearly polarized along a certain direction in the film called the transmission axis (*hatching*). This polarization state of the photon is represented by the wavy colored line at the top. The film blocks all light incident on it at a right angle if the light is linearly polarized perpendicular to the transmission axis (*wavy gray line at top*). Now suppose a photon is linearly polarized at some angle to the transmission axis between zero and 90 degrees (*bottom*). Then whether or not the photon will be transmitted is indefinite; the probability of transmission is a number between zero and 1 (the square of the cosine of the angle).

the horizontal: either both photons will be transmitted or both will be blocked. It simply cannot happen that one photon will be transmitted and the other will be blocked. In fact, it does not matter what the orientations of the films are as long as they match each other; the outcomes of the linear-polarization experiments are strictly correlated for an infinite family of possible experiments. (Of course, no more than one of the experiments can actually be carried out.) Somehow the second photon of the pair "knows" whether to pass through its polarizing film in order to agree with the passage or nonpassage of the first photon, even though the two photons are well separated and neither has a mechanism for informing the other of its behavior. In this kind of situation, then, quantum mechanics challenges the relativistic concept of locality, which holds that an event cannot have effects that propagate faster than light (and, in particular, instantaneous effects at a distance).

It must be emphasized that all the peculiar implications that have been drawn so far—objective indefiniteness, objective chance, objective probability and nonlocality—depend crucially on the premise that a system's quantum state is a complete description of that system. A number of theorists have maintained, however, that the quantum state merely describes an ensemble of systems prepared in a uniform manner, and that this is why good predictions can be made about the statistical results of the same experiment performed on all members of the ensemble. At the same time, the argument goes, the individual members of the ensemble differ from one another in ways not mentioned by the quantum state, and this is the reason the outcomes of the individual experiments are different. The properties of individual systems that are not specified by the quantum state are known as hidden variables.

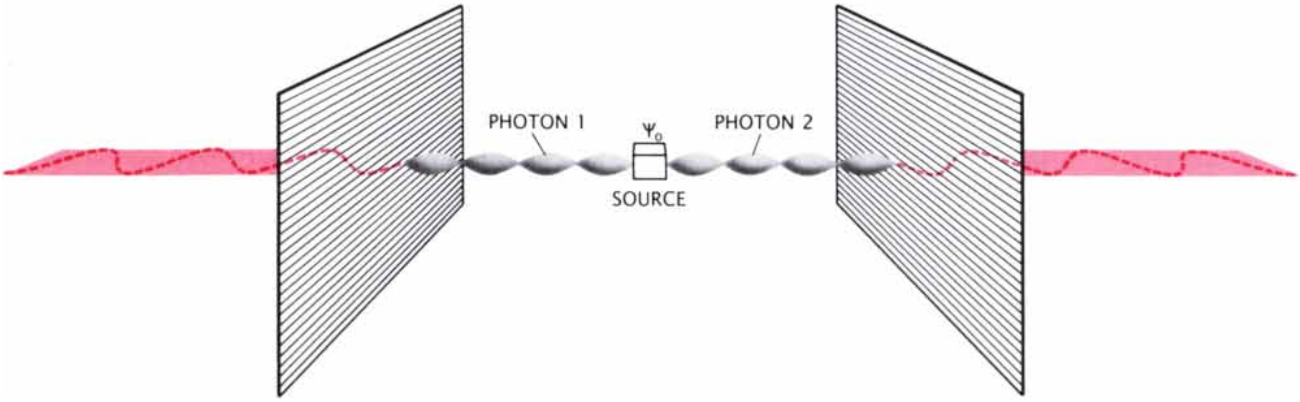
If hidden-variables theorists are correct, there is no objective indefiniteness. There is only ignorance on the part of the scientist about the values of the hidden variables that characterize an individual system of interest. Moreover, there is no objective chance and there are no objective probabilities. Most important, the quantum correlations of well-separated systems are no more surprising than the concordance of two newspapers printed by the same press and mailed to different cities.

In 1964 John S. Bell of CERN, the European laboratory for particle physics, showed that the predictions of local hidden-variables models are incompatible with the predictions of quantum mechanics. Reflection on some hidden-variables models of David Bohm of Birkbeck College London and Louis de Broglie led Bell to prove the important theorem that no model that is local (in a carefully specified sense) can agree with all the statistical predictions of quantum mechanics. In other words, there are physical situations in which the predictions of quantum mechanics disagree with those of every local hidden-variables model [see "The Quantum Theory and Reality," by Bernard d'Espagnat; SCIENTIFIC AMERICAN, November, 1979].

The idea of Bell's theorem can be grasped, at least in part, by returning to consider the quantum state Ψ_0. As noted above, the results of linear-polarization experiments performed on a pair of photons in this state must be strictly correlated when the angle between the transmission axes of the two polarizing films is zero degrees (as it is when both axes are aligned vertically). It should not be surprising to learn, therefore, that for the state Ψ_0 there is always at least a partial correlation between the outcomes, no matter what the angle between the transmission axes is. (Specifically, if one of the photons is transmitted through its polarizing film, then the probability that the other photon will be transmitted through its film is the square of the cosine of the angle between the two transmission axes.)

Consequently a hidden-variables model that agrees with all the statistical predictions of quantum mechanics must assign quantities to each pair of photons in the ensemble in a delicate way in order to guarantee the strict or partial correlations at every angle between the axes. But the condition of locality requires that the quantities assigned to each photon in a pair must be independent of the orientation of the polarizing film on which the other photon impinges and independent of the other photon's passage or nonpassage. It is this locality condition that makes quite impossible the delicate adjustments that would be necessary for reproducing all the correlations, strict and partial, implied by Ψ_0.

Bell's theorem shows that in principle one can determine experimentally which is correct: quantum mechanics or the local hidden-variables models. It was important to do such a test because, in spite of the immense body of confirming evidence



CORRELATIONS between the polarizations of two photons occur when the photons are in a special state called Ψ_0 (after the letter psi in the Greek alphabet). The state can be formed by superposing the state in which both photons are linearly polarized along a vertical axis with the state in which they are both linearly polarized along a horizontal axis. The state Ψ_0 contains equal amounts of the vertically polarized state and the horizontally polarized state. Now imagine that polarizing films with horizontally oriented transmission axes are inserted in the paths of the photons. Since Ψ_0 contains equal amounts of the two states, there is a 50 percent probability that both photons will be transmitted through their respective films and a 50 percent probability that both will be blocked. What cannot happen is that one photon will be transmitted and the other will be blocked: the outcomes of the linear-polarization experiments are strictly correlated. In fact, it does not matter what the orientations of the films are as long as they match each other; somehow the second photon of the pair "knows" whether to pass through its polarizing film in order to agree with the passage or nonpassage of the first photon, even though the photons are well separated.

for quantum mechanics at the time Bell proved his theorem, the very points where quantum mechanics is without equivocation irreconcilable with common sense had not yet been probed.

In 1969 John F. Clauser, then at Columbia University, Michael A. Horne of Boston University, Richard A. Holt, then at Harvard University, and I proposed a design for the requisite test. Pairs of photons with correlated linear polarizations were to be obtained by exciting atoms to an appropriate initial state; the atoms would subsequently return to the unexcited state by emitting two photons. Filters and lenses would ensure that when the photons flew off in opposite or virtually opposite directions, one photon would impinge on a polarization analyzer and the other would impinge on another analyzer. By switching between two orientations of each analyzer and recording the number of photon pairs transmitted in each of the four possible combinations of orientations of the two analyzers, measurements of correlations of transmissions of the photons of a pair could be made.

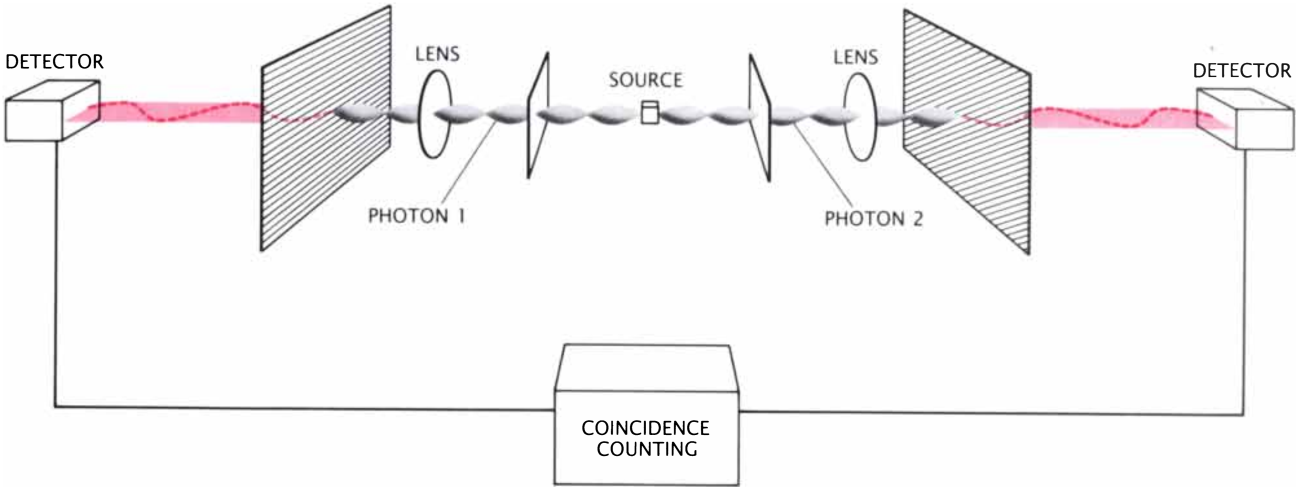
We suggested that either calcite crystals or piles of glass plates serve as the polarization analyzers, since each of them is much more efficient than an actual polarizing film in blocking photons polarized perpendicular to the transmission axis. Photodetectors placed behind the analyzers would detect a fixed fraction of the photons passing through the analyzers. If two photons, one at each detector, were registered within 20 nanoseconds (billionths of a second) of each other, the probability would be quite high that they were emitted by the same atom. Since the lenses would collect the two photons over a finite angle, the quantum state would not be exactly the Ψ_0 state discussed above but a modified state Ψ_1, which also leads to correlations that cannot be reproduced by any local hidden-variables model.

The experiment was done by Stuart J. Freedman and Clauser at the University of California at Berkeley in 1972, by Edward S. Fry and Randall C. Thompson at Texas A. & M. University in 1975 and by other groups after that. Most of the experimental results agree with the correlation predictions of quantum mechanics and disagree with the hidden-variables models. Moreover, the reliability of the dissenting experiments is suspect because of subtle weaknesses in their design.

Yet until very recently all the experiments had a loophole that allowed staunch defenders of hidden-variables models to maintain their hopes: the polarization analyzers were kept in their respective orientations for intervals of a minute or so, which is ample time for the exchange of information between the analyzers by some hypothetical mechanism. As a result the defenders could contend that the special theory of relativity does not imply the validity of Bell's locality condition in the physical situation of the experiments. But then these experiments would not serve as decisive tests between quantum mechanics and the local hidden-variables models.

To block this loophole, Alain Aspect, Jean Dalibard and Gérard Roger of the Institute of Optics of the University of Paris did a spectacular experiment in which the choice between the orientations of the polarization analyzers is made by optical switches while the photons are in flight. In their experiment, which required eight years of work and was completed only in 1982, each switch is a small vial of water in which standing waves are periodically generated ultrasonically. The waves serve as diffraction gratings that can deflect an incident photon with high efficiency. If the standing waves are turned on, the photon will be deflected to an analyzer that is oriented one way; if the standing waves are turned off, the photon will travel straight to an analyzer that is oriented another way.

The switching between the orientations takes about 10 nanoseconds. The generators that power the two switches operate independently, although (unfortunately for the complete definitiveness of the experiment) the operation is periodic rather than random. The distance between the analyzers is 13 meters, so that a signal moving at the speed of light (the highest speed allowed by the special theory of relativity) takes 40 nanoseconds to travel between them. Consequently the choice of orientation for the first polarization



SEARCH FOR CORRELATIONS between members of pairs of photons was carried out in the 1970's by a number of investigators. The photon pairs were emitted in energy-state transitions of calcium and mercury atoms; each photon impinged on a polarization analyzer. Quantum mechanics predicts there must be delicate correlations in the passage or nonpassage of the photons through their analyzers, even though the photons have no apparent means of communicating with each other. The experiments mainly confirmed quantum mechanics, but they had a loophole: the orientations of the two analyzers were fixed before the photons were emitted. Consequently it was possible that information was somehow exchanged between the analyzers.

analyzer ought not to influence the transmission of the second photon through the second analyzer, and the choice of orientation for the second analyzer ought not to influence the transmission of the first photon through the first analyzer. The experimental arrangement is thus expected to satisfy Bell's locality condition. It follows that—according to Bell's theorem—there should be some violations of the quantum-mechanical predictions of correlations in the experimental outcome.

In point of fact, however, the experiment yielded just the opposite result. The correlation data agree within experimental error with the quantum-mechanical predictions that are calculated on the basis of the quantum state Ψ_1. Moreover, the data disagree by more than five standard deviations with the extreme limits allowed, according to Bell's theorem, by any of the local hidden-variables models.

Even though the experiment of Aspect and his colleagues is not completely definitive, most people believe the prospects of overthrowing the results by future experiments are extremely small. It seems unlikely that the family of local hidden-variables models can be salvaged. The strange properties of the quantum world—objective indefiniteness, objective chance, objective probability and nonlocality—would appear to be permanently entrenched in physical theory.

One of the strangest properties of the quantum world is nonlocality. Can the fact that under some circumstances a measurement on one photon apparently instantaneously affects the result of a measurement on another photon be capitalized on to send a message faster than the speed of light? Fortunately for the special theory of relativity the answer to the question is no. An underlying assumption of that theory—that no signal can travel faster than light—is preserved.

Here is a brief argument that shows why. Suppose two people want to communicate by means of a device similar to the one for testing local hidden-variables models. Between the observers a source emits pairs of correlated photons. Each observer is provided with a polarization analyzer and a photodetector. The observers are free to orient the transmission axes of their analyzers any way they choose.

Suppose the observers agree to align the transmission axes vertically. Then every time a pair of photons is emitted there will be a strict correlation in the outcome: either both photons will pass through the analyzers or both will be blocked. But the strict correlation is of no value for each observer in isolation from the other. The first observer will note that half of the time photons pass through the first analyzer, on the average, and half of the time they are blocked. The second observer will note the same thing for the second analyzer. In other words, each observer in isolation sees only a random pattern of transmissions and blockages.

Now imagine that the first observer tries to encode some information and send it to the second observer by changing the orientation of the first polarization analyzer. Depending on the orientation of that analyzer, there will be either a strict or a partial correlation between the outcomes of the events at each detector. Once again, however, each observer will note that on the average half of the time photons pass through the analyzer and half of the time they are blocked. In general, no matter what the orientations of the analyzers are, each observer in isolation sees only a random (and statistically identical) pattern of transmissions and blockages. The quantum correlations between the photons can be checked only by comparing the data accumulated at the two detectors. Hence the attempt to exploit the quantum correlations to send messages faster than light cannot succeed.

In this sense there is a peaceful coexistence between quantum mechanics and relativity theory, in spite of quantum-mechanical nonlocality. For this reason it would be misleading (and wrong) to say that nonlocality in the quantum-mechanical sense is a reversion to action at a distance, as in the prerelativistic gravitational theory of Newton. It is tempting to characterize quantum-mechanical nonlocality as "passion at a distance," not with any pretension to provide an explanation for the strange correlations, but only to emphasize that the correlations cannot be exploited to exert a controlled influence more rapidly than a light signal can be sent.

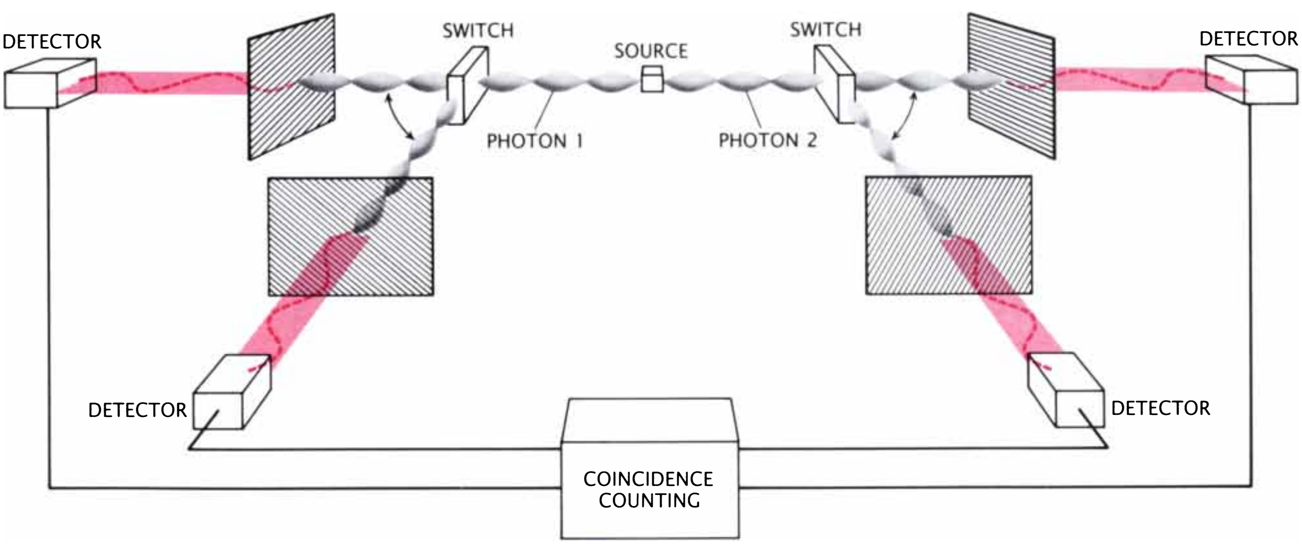
Another test, called the delayed-choice experiment, which was proposed in 1978 by John Archibald Wheeler, then at Princeton University, also reveals the strangeness of the quantum world. The basic apparatus of the experiment is an interferometer, in which a light beam can be split and recombined. A pulse of light from a laser is fired at the beam splitter, which is oriented in such a way that half of the light passes through the splitter and half is reflected at right angles to the direction of the incident pulse. If the light from the two paths is subsequently recombined, an interference pattern can be detected, which demonstrates the wavelike quality of light.

Now suppose the pulse of laser light is attenuated so severely that at any time there is only one photon in the interferometer. In this case two different questions can be asked about the photon. Does the photon take a definite route so that it is either transmitted or reflected by the beam splitter, thereby exhibiting a particlelike property? Or is the photon in some sense simultaneously transmitted and reflected so that it interferes with itself, thereby showing a wavelike property?

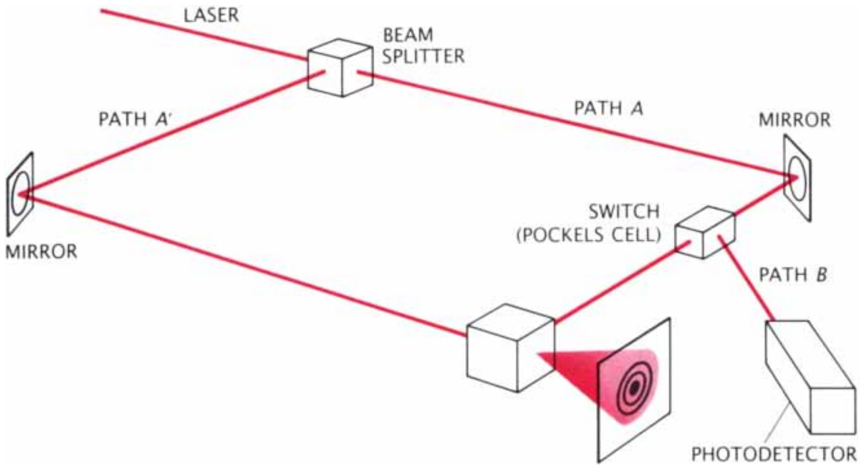
An answer was recently supplied by Carroll O. Alley, Oleg G. Jakubowicz and William C. Wickes of the University of Maryland at College Park and independently by T. Hellmuth, H. Walther and Arthur G. Zajonc of the University of Munich. Both groups found that a photon behaves like a particle when particlelike properties are measured and that it behaves like a wave when wavelike properties are measured. The remarkable novelty of the results is that the experiment was arranged so that the decision to measure particlelike or wavelike properties was made after each photon had interacted with the beam splitter. Consequently the photon could not have been "informed" at the crucial moment of interaction with the beam splitter whether to behave like a particle and take a definite route or to behave like a wave and propagate along two routes.

The length of both routes in the interferometer was about 4.3 meters, which a photon can traverse in roughly 14.5 nanoseconds. Obviously this does not allow enough time for an ordinary mechanical device to switch between measuring particle- and wavelike properties. The feat was made possible with a switch called a Pockels cell, which can be actuated in nine nanoseconds or less. A Pockels cell contains a crystal that becomes birefringent when a voltage is applied across it: light polarized along one axis of the crystal propagates at a velocity different from that of light polarized along the perpendicular direction. Given the proper choice of voltage and configurational geometry, light polarized in one direction when it enters the cell will emerge polarized in the perpendicular direction. The Pockels cell was inserted in one of the two routes the photon could take after interacting with the beam splitter [*see illustration on next page*].

A piece of polarizing film was the other essential element needed to switch between measurements of particlelike and wavelike properties. Light emerging from the Pockels cell impinged on the film. If the cell was "on," the polarization of the light was such that the polarizing film reflected the light into a photodetector, thereby answering the question of which route and confirming the photon's particlelike properties. If the cell was "off," the polarization of the light was such that the polarizing film transmitted the light, which then was combined with the contribution coming from the other route; interference ef-



RAPID SWITCHING between orientations of polarization analyzers as photons flew was the hallmark of the experiment done by Aspect and his colleagues (*see illustration on page 47*), which was completed in 1982. When a switch was "on," a photon was deflected to an analyzer that was oriented one way; when the switch was "off," the photon traveled straight to an analyzer that was oriented another way. The time required for light to travel between the analyzers was greater than the time required to switch between orientations, so that the choice of orientation for each analyzer could not influence the observation made at the other analyzer. (Unfortunately for complete definitiveness, however, the switching was periodic rather than random.) The experiment confirmed quantum mechanics; it would appear that the strange implications of the theory must be accepted.



DELAYED-CHOICE EXPERIMENT is another test that reveals the strangeness of the quantum world. A photon impinges on a beam splitter. Two questions about the photon can be asked. Does the photon take a definite route so that it is either transmitted or reflected by the beam splitter, thereby exhibiting a particlelike property? Or is the photon in some sense both transmitted and reflected so that it interferes with itself, exhibiting a wavelike property? To find out, a switch is positioned in one of the two paths the photon can take after interacting with the beam splitter (*here, path* A). If the switch is on, the light is deflected into a photodetector (*path* B), thereby answering the question of which route and confirming the photon's particlelike properties. If the switch is off, the photon is free to interfere with itself (*paths* A *and* A') and produce an interference pattern, demonstrating the photon's wavelike properties. Results from the experiment show that a photon behaves like a wave when wavelike properties are measured and behaves like a particle when particlelike properties are measured. Remarkably, the switch was triggered after the photon had interacted with the beam splitter, so that the photon could not have been "informed" whether to behave like a particle and take a definite route or to behave like a wave and propagate simultaneously along two routes.

fects confirmed the photon's wavelike aspect.

Both groups of investigators have reported results that are in excellent agreement with quantum mechanics. Their work shows that the choice between the two questions can be made after an individual photon has interacted with the beam splitter of an interferometer.

How are the results of the delayed-choice experiment to be interpreted? It is worthwhile first to disclaim one extravagant interpretation that has sometimes been advanced: that quantum mechanics allows a kind of "reaching into the past." Quantum mechanics does not cause something to happen that had not happened previously. Specifically, in the delayed-choice experiment quantum mechanics does not cause the photon to take a definite route at time zero if 12 nanoseconds later the Pockels-cell switch is turned on, and it does not cause the photon to take both routes, in wavelike fashion, if the switch is off.

A more natural interpretation is that the objective state of the photon in the interferometer leaves many properties indefinite. If the quantum state gives a complete account of the photon, then that conclusion is not surprising, since in every quantum state there are properties that are indefinite. But the conclusion does raise a further question: How and when does an indefinite property become definite? Wheeler's answer is that "no elementary quantum phenomenon is a phenomenon until it is a registered phenomenon." In other words, the transition from indefiniteness to definiteness is not complete until an "irreversible act of amplification" occurs, such as the blackening of a grain of photographic emulsion. Students of the foundations of quantum mechanics disagree about the adequacy of Wheeler's answer, however. The next experiment shows why the question is still open.

In 1935 Erwin Schrödinger proposed a famous thought experiment. A photon impinges on a half-silvered mirror. The photon has a probability of one-half of passing through the mirror and a probability of one-half of being reflected. If the photon passes through the mirror, it is detected, and the detection actuates a device that breaks a bottle of cyanide, which in turn kills a cat in a box. It cannot be determined whether the cat is dead or alive until the box is opened.

There would be nothing paradoxical in this state of affairs if the passage of the photon through the mirror were objectively definite but merely unknown prior to observation. The passage of the photon is, however, objectively indefinite. Hence the breaking of the bottle is objectively indefinite, and so is the aliveness of the cat. In other words, the cat is suspended between life and death until it is observed. The conclusion is paradoxical, but at least it concerns only the results of a thought experiment.

It is now more difficult to dismiss the paradoxical nature of the conclusion, because something similar to Schrödinger's thought experiment has recently been achieved by a number of groups of investigators including Richard F. Voss and Richard A. Webb of the IBM Thomas J. Watson Research Center in Yorktown Heights, Lawrence D. Jackel of the AT&T Bell Laboratories, Michael H. Devoret of Berkeley and Daniel B. Schwartz of the State University of New York at Stony Brook. Their work has relied to a certain extent on calculations that were done by Anthony J. Leggett of the University of Illinois at Urbana-Champaign and Sudip Chakravarty at Stony Brook, among other investigators.

The experimental apparatus consists of an almost closed superconducting ring. A thin slice of insulating material (called a Josephson junction) interrupts the ring, but an electric current can circulate around the ring by "tunneling" through the insulator. The current generates a magnetic field.

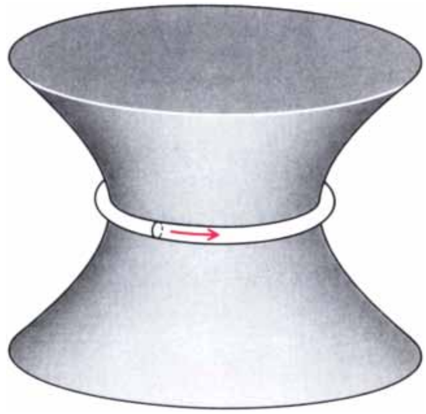
The quantity that is of interest in the system is the magnetic flux through the ring, which (when the field is uniform) is equal to the area of the ring multiplied by the component of the magnetic field perpendicular to the plane of the ring. If the ring were uninterrupted, the flux would be trapped within the ring, but the insulator allows the flux to slip from one value to another. With modern magnetometers the flux can be measured with fantastic accuracy. The fact that the flux arises from the motion of enormous numbers of electrons (on the order of 10^{23}) in the superconducting ring justifies speaking of the flux as a macroscopic quantity. There is now good evidence that states of the superconducting ring can be prepared in which the flux does not have a definite value—a quantum-mechanical feature that had previously been established only for observables of microscopic systems.

To understand how this indefiniteness is demonstrated experimentally, it is necessary to know that for each value of the flux the ring has a certain potential energy. Ordinarily one would not expect that the flux through the ring could change spontaneously from one value to another, because a potential-energy barrier separates adjacent values of the flux from each other. Classical physics forbids the transition between two such values of the flux unless some external source of energy, typically thermal, is supplied to traverse the barrier between them. In quantum mechanics, on the other hand, the barrier can be tunneled through without any external source of energy. The groups of investigators mentioned above have shown that the flux does change between two values, and that the change cannot be entirely accounted for thermally; the observed tunneling must be at least partially quantum mechanical, particularly at very low temperatures. But quantum-mechanical tunneling rests essentially on the indefiniteness of the flux, which thus cannot be localized definitely at or close to one value or another.

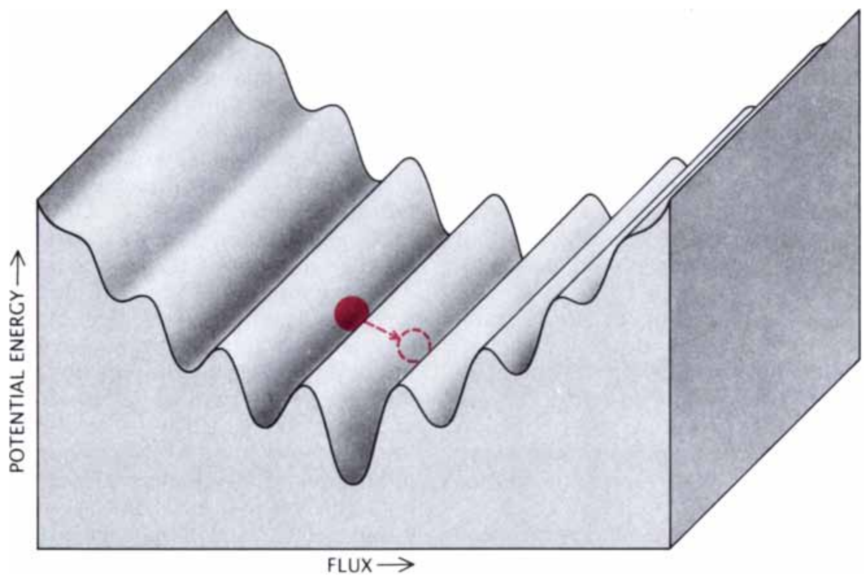
The experimental demonstration of quantum indefiniteness in a macroscopic variable does not ipso facto contradict the statement by Wheeler quoted above, but it does show that amplification from a microscopic to a macroscopic level does not in itself exorcise quantum-mechanical indefiniteness. The emphasis in Wheeler's statement about an "irreversible act of amplification" must be placed on the word "irreversible." The conditions for the occurrence of an irreversible process are far from settled in contemporary theoretical physics. Some students of the subject (including me) believe new physical principles must be discovered before we can understand the peculiar kind of irreversibility that occurs when an indefinite observable becomes definite in the course of a measurement.

The strangeness of the quantum world continues to be explored. Still other experiments have recently been performed or are now under way; two classes of these experiments should be mentioned here, even though there is no room to discuss them in detail. In the neutron-interferometer experiments of Helmut Rauch and Anton Zeilinger of the Atomic Institute of the Austrian Universities, Samuel A. Werner of the University of Missouri at Columbia and Clifford G. Shull of the Massachusetts Institute of Technology and their associates, the wave function of a neutron is split by a sheet of crystal and recombined by one or two other sheets. The interference effects exhibited in the recombination demonstrate a number of remarkable properties, including the indefiniteness of the neutron's route through the interferometer.

Finally, R. G. Chambers of the University of Bristol, G. Möllenstedt of the University of Tübingen and Akira Tonomura of Hitachi, Ltd., have confirmed by electron interferometry the remarkable Aharonov-Bohm effect, in which an electron "feels" the presence of a magnetic field that is in a region where there is zero probability of finding the electron. This is a striking demonstration of a kind of nonlocality different from, although remotely related to, the nonlocality exhibited by correlated photon pairs. A thorough understanding of the relation between the two kinds of nonlocality as well as the many other perplexing issues raised by experiments probing the nature of the quantum world awaits further work.



MACROSCOPIC SYSTEM can under some circumstances exist in a state in which a macroscopic variable has an indefinite value; indefiniteness is not limited to microscopic systems, such as the photon. The system shown here is a superconducting ring that does not quite bend back on itself. A thin slice of insulating material separates the two ends of the ring from each other, and an electric current is made to circulate around the ring by "tunneling" through the insulator. The current generates a magnetic field. If the ring were continuous, the magnetic flux through the ring (the area of the ring multiplied by the component of the magnetic field perpendicular to the plane of the ring) would be trapped at a fixed value, but the insulator allows the flux to slip from one value to another. Surprisingly, the flux does not have a definite value.



INDEFINITENESS in the system shown at the top of the page is depicted schematically. Each value of the flux through the superconducting ring has a certain potential energy associated with it. Ordinarily one would not expect that the flux through the ring could spontaneously change from one value to another, because a potential-energy barrier separates the adjacent values of the flux from each other. The barriers can be thought of as hills, and the state the system is in can be represented as a ball residing in a valley among the hills. According to classical physics, a transition between two values separated by a barrier requires outside energy (to push the ball over the hill). Quantum mechanically, however, the barrier can be tunneled through without any external source of energy. Tunneling is essentially a manifestation of the indefiniteness of the flux.

The Very-Long-Baseline Array

An array of 10 radio antennas across the U.S. will provide the most detailed images yet of the universe. With it astronomers will explore such cosmic puzzles as the mysterious processes powering the quasars

by Kenneth I. Kellermann and A. Richard Thompson

Ever since Galileo pointed his telescope to the night sky nearly 400 years ago, astronomers have been building instruments of ever increasing sophistication with which to observe the universe. Each improvement in the resolving power of their instruments has enabled them to examine the universe in increasing detail and to discover new kinds of objects unknown to earlier generations. Galileo's telescope achieved a twentyfold improvement in resolution and allowed human eyes to see for the first time the phases of Venus, the rings around Saturn, the four bright moons of Jupiter, the craters and mountains on the moon and the myriad stars of the Milky Way. The giant optical instruments of today, such as the 200-inch Hale telescope on Mount Palomar, can detect objects more than a million times fainter than those Galileo could see. But because of limits imposed by turbulence in the atmosphere, they can distinguish features only one-tenth the size of those that could be detected with Galileo's rudimentary telescope.

The development of radio technology in World War II opened a completely new window on the universe. When astronomers turned radio antennas to the heavens, they began to find a previously unknown universe of solar and planetary radio bursts, quasars, pulsars, radio galaxies, giant molecular clouds and cosmic masers. Not only do the radio waves reveal a new world of astronomical phenomena but also—because they are much longer than light waves—they are not as severely distorted by atmospheric turbulence or small imperfections in the telescope.

To the pioneers in radio astronomy, however, the long wavelengths appeared to be a severe handicap. The resolving power of a telescope depends on the wavelength divided by the diameter of the aperture. To obtain a resolution comparable to that of an optical telescope operating at a typical light wavelength of 5,000 angstrom units (5×10^{-7} meter), a radio antenna operating at a one-meter wavelength would have to be a million times as large. Therefore although the early radio telescopes could detect signals from remote galaxies that are faint or invisible with even the largest optical telescopes, they had such poor resolution that they could not always tell individual sources apart. Even the largest steerable single-dish antenna, a 100-meter reflector in Germany operating at wavelengths of about one centimeter, can attain a resolution of only one minute of arc, roughly the same as that of the unaided eye. To build a radio telescope with a resolution of one second of arc, comparable to that of the Hale telescope, would require an antenna with a diameter in the tens of kilometers.

Fortunately there is a way around the dilemma. About 25 years ago radio astronomers became aware that they could synthesize a resolution equivalent to that of a large aperture by combining data from smaller radio antennas that are widely separated. The effective aperture size would be about equal to the largest separation between the antennas. The technique is called synthesis imaging and is based on the principles of interferometry. Radio astronomers in the U.S. are now building a synthesis radio telescope called the Very-Long-Baseline Array, or VLBA. With 10 antennas sited across the country from the Virgin Islands to Hawaii, it will synthesize a radio antenna 8,000 kilometers across, nearly the diameter of the earth. The VLBA's angular resolution will be less than a thousandth of an arc-second—about three orders of magnitude better than that of the largest conventional ground-based optical telescopes. Astronomers eagerly await the completion early in the next decade of the VLBA, which is expected, among other things, to give an unprecedentedly clear view into the cores of quasars and galactic nuclei and to reveal details of the processes—thought to be powered by black holes—that drive them.

Radio Interferometry

The basic principle of the VLBA is that of the radio interferometer, a system that combines the signals received from a radio source by two or more antennas. The resulting interference pattern indicates the difference in the path lengths from the source to the antennas. If the paths differ by an integral number of wavelengths, wave crests will arrive at the antennas in phase, or simultaneously, and produce an intensity maximum when the signals are combined. Conversely, if the paths differ by an odd number of half wavelengths, a crest and a trough will coincide and cancel, thereby creating an intensity minimum in the combined signal. Because the path lengths are determined by the position of the radio source in relation to the antennas, the interference pattern contains information about the location and detailed features of the source and can be used to construct an image.

When a celestial object is observed with a radio interferometer, the rotation of the earth causes the path-length difference to vary and the received signal to oscillate between being in and out of phase, creating a sinusoidal pattern of maximums and minimums called interference fringes. The earth must rotate through a greater angle between one maximum and the next for short baselines than it must for long ones. Closely spaced antennas therefore produce broad fringes and respond only to the coarser structure of a source. Antennas with wider spacings respond to finer detail. To gather full information on the structure of a source it is necessary to have an array of antennas with a variety of baseline lengths. Moreover, the baseline orientations must be carefully distributed in order to obtain good two-dimensional images. The rotation of the earth itself causes foreshortening and reorientation of a baseline in relation to the source, so that a series of observations over a period of time has the effect of adding more baselines to the array.

The most powerful radio telescope operating today, the Very Large Array (VLA), is a giant Y-shaped array spread out on a high desert plateau in central New Mexico. It consists of 27 fully steerable parabolic antennas distributed along the three arms of the Y. A 21-kilometer railroad track runs along each arm with fixed stations along the way that provide bases for the antennas. The antennas are connected by buried waveguides to a central facility where the incoming signals are combined to produce the interference fringes. The nine antennas on each arm can be moved along the track and arranged in four different configurations so that the arm lengths range from 600 meters to 21 kilometers. The antennas operate in several frequency bands from 330 megahertz (90-centimeter wavelength) to 23 gigahertz (1.3-centimeter wavelength). The VLA yields high-sensitivity images with a resolution as good as a tenth of an arc-second, and the image quality equals or surpasses that of the best ground-based optical telescopes. With the VLA astronomers have obtained radio images of such objects as sunspots, the rings of Saturn, dark clouds in our galaxy and the mysterious energetic jets emanating from quasars and from the center of radio galaxies.

But many cosmic radio sources, such as the enigmatic quasars, are much too small to be resolved even with the VLA. They can be studied only by extending the antenna spacings to thousands of kilometers. Because of the distances involved, the antennas cannot be physically interconnected. Instead signals received at each antenna are recorded on magnetic tapes, which are transported to a central laboratory and there replayed simultaneously to simulate the resolution of a single enormous antenna. This technique is called very-long-baseline interferometry (VLBI), and it has been exploited with great success since the 1960's [see "Radio Astronomy by Very-Long-Baseline Interferometry," by Anthony C. S. Readhead; SCIENTIFIC AMERICAN, June, 1982]. Every few months radio observatories around the globe, mainly in the U.S. and Europe but including at times antennas in Australia, Brazil, Canada, China, India, Japan, South Africa and the Soviet Union, coordinate their schedules to track selected objects. Tapes recorded at each of the antennas are then replayed at one of three processing locations: the Max Planck Institute for Radio Astronomy in Bonn, the National Radio Astronomy Observatory in Charlottesville, Va., and the California Institute of Technology. As many as 18 antennas have taken part in this VLBI system to obtain remarkable images of quasars, active galactic nuclei, cosmic masers and other compact radio sources.

Details of the VLBA

This ad hoc VLBI system nonetheless leaves much to be desired. It is not easy to arrange for a sufficient amount of coordinated observ-

VERY-LONG-BASELINE ARRAY (VLBA) of 10 radio antennas will extend 8,000 kilometers across the U.S. and have a resolution comparable to that of a single antenna nearly as wide as the earth. Each antenna will record radio signals from cosmic sources, and the data will be correlated at a central facility to generate interference patterns from which images can be obtained. The Y-shaped symbol marks the location of the Very Large Array (VLA), which will provide additional data to supplement the VLBA.

Festiva. The product of Ford's worldwide resources.

A direct result of Ford's worldwide resources and engineering expertise, Festiva was first introduced in Japan. It was a sales success. And a design award winner.

Little wonder. It's a car that offers quick, maneuverable fun.

Festiva. The fun starts at $5490.

And it includes front-wheel drive. A 1.3 liter 4-cylinder engine. Power front-disc brakes. Rack and pinion steering. High-back cloth bucket seats. Flip fold-down rear seats. Side window demisters. All standard.

Or you may choose Festiva LX, which offers things like a 5-speed manual transaxle. Tilt steering column. Electronic AM/FM stereo. Dual electric remote control mirrors. Rear window wiper and defroster. And more. All for **$6,801 as shown.** (Sticker prices exclude title, destination charges and taxes. Aluminum wheels shown, $396. Prices higher in NY.)

Festiva. A world of room inside.

Festiva's hatchback design rates an interior volume index of 98.4 cubic feet. Which means

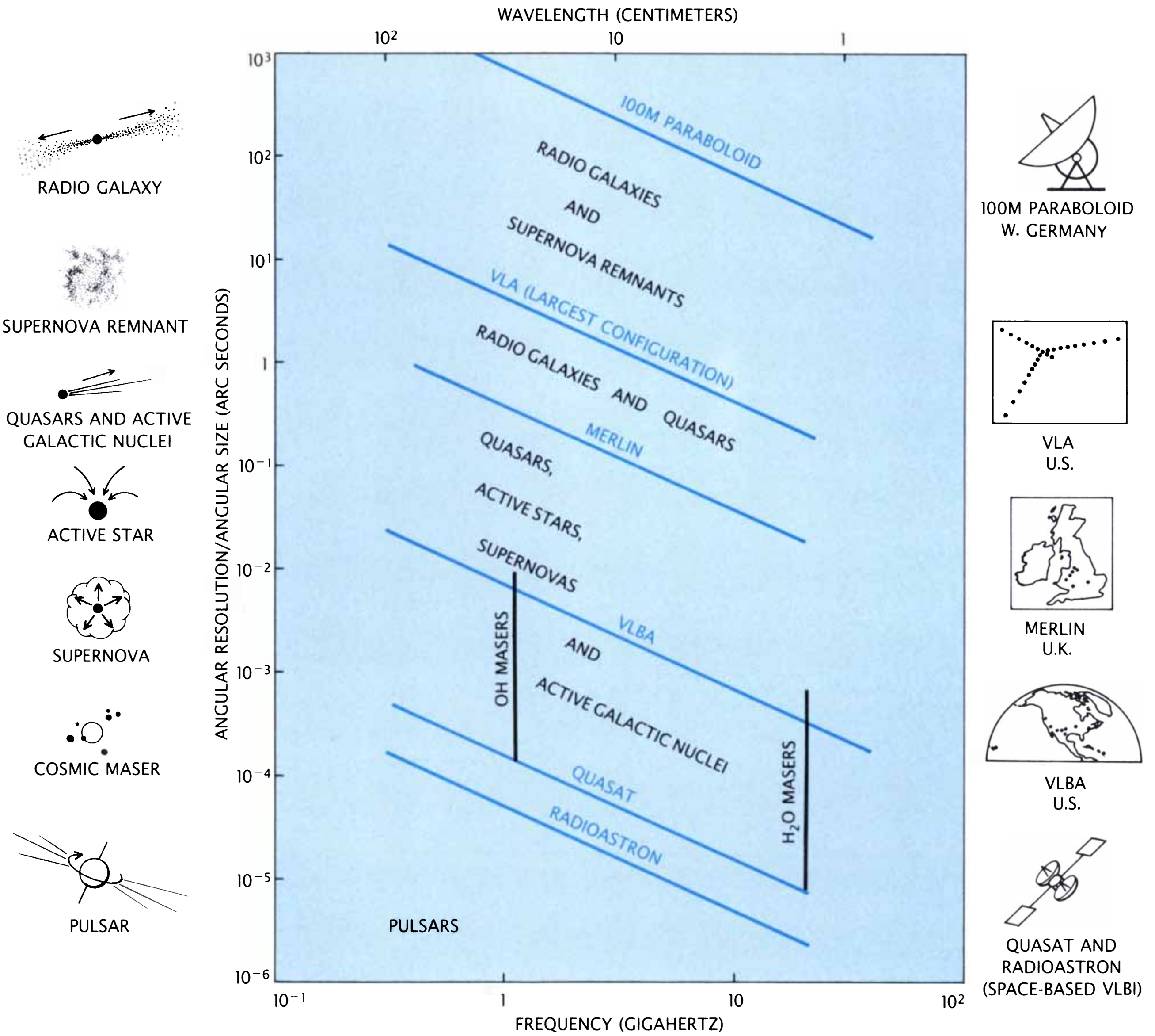
Welcome a new kind of Ford to the street. Welcome Ford Festiva.


Festiva LX

ing time, and individual antennas are not properly sited to give good image quality. Moreover, the antennas vary in accuracy and sensitivity. In order to exploit the VLBI technique better, the National Radio Astronomy Observatory, with funds from the National Science Foundation, began in 1985 to construct the VLBA, a dedicated VLBI network of 10 antennas. Each antenna will be 25 meters in diameter and designed to operate at wavelengths as short as 3.5 millimeters. The antennas will be at sites throughout the U.S. chosen so that the distribution of baselines will provide good image quality. Other important objectives in the choice of antenna sites were to avoid strong manmade radio signals, to minimize atmospheric water vapor, to be near major transportation centers and to have access to local technical support. Five of the antennas will be at high elevations in the relatively dry, cloudless southwestern states.

Each antenna will be controlled through a local computer that will receive instructions over telephone lines from the VLBA Operations Center in Socorro, N.M. The array will follow a planned program under the control of a central computer that will monitor the antennas and receivers as well as weather conditions at each site. An array-control operator will be able to make changes on short notice if there is an unexpected cosmic event, such as a supernova. The operator will also be able to intervene when necessary, for example if there are technical problems or poor atmospheric conditions. For special experiments requiring even better images it will be possible to include data from as many as 10 other radio telescopes around the world.

At the present time the first VLBA



RESOLUTION OF RADIO TELESCOPES is plotted against the frequency of observed radio emissions for various antennas and arrays. The larger the effective aperture, the finer the resolution (measured by the angle subtended by the smallest distinguishable feature). The graph also indicates the angular size of various types of astronomical radio sources. The Bonn telescope is a single, 100-meter antenna. The Very Large Array (VLA) consists of 27 movable antennas physically linked by waveguide. MERLIN, operated by the Nuffield Radio Astronomy Laboratories in England, is a seven-antenna array extending over 200 kilometers and linked by microwave transmission. The VLBA, with a maximum baseline of 8,000 kilometers, will use recorded data. *Quasat* and *Radioastron* orbiting telescopes will employ both microwave links and recordings to extend baselines into space.

antenna, in Pie Town, N.M., is coming into operation, and five others are in various stages of construction. The full VLBA should be completed by 1992. By the mid-1990's astronomers hope to see the first antenna for radio interferometry launched into space. The VLBA will at that time provide ground-based observations to be combined with the data gathered from space. The VLBA, then, is not only a major advance in high-resolution astronomy but also a step toward the even more powerful space-based arrays of the future.

Clocks and Recorders

Fifty years of innovations, from hydrogen-maser clocks to home videocassette recorders, have contributed to the technologies that make the VLBA possible. The highly accurate clocks are needed to synchronize the data from the antennas; the magnetic-tape recorders store the huge volumes of data collected at each antenna. The radio receivers at each antenna will be among the most sensitive available. Most of the receivers will have transistor preamplifiers that will be cooled to 15 degrees Kelvin (degrees Celsius above absolute zero) to minimize noise in the instruments. Each antenna will be equipped to cover nine separate bands in the frequency range from 330 megahertz (90-centimeter wavelength) to 43 gigahertz (seven-millimeter wavelength). In the future the frequency range may be extended to 86 gigahertz (3.5-millimeter wavelength), which approaches the operational limit of the antennas.

In order for the VLBI technique to work there must be an accurate clock at each antenna so that the data can be synchronized. In addition, because the data are converted into a lower frequency, there must be an extremely stable frequency standard so that the phase of the signal is preserved. Both functions will be carried out in the VLBA by a hydrogen-maser clock, which bases its time standard on a characteristic frequency of the hydrogen atom. (Maser stands for microwave amplification by stimulated emission of radiation—the microwave counterpart of the laser.) The clock's frequency is stable to a few parts in 10^{15} for periods of up to an hour. This means that at the VLBA's highest frequency of 43 gigahertz the data from separate antennas can be synchronized over a time interval of about half an hour before the relative phases drift by an appreciable amount. It is therefore possible to condense half-hour segments of data by taking the average and thus to greatly reduce the amount of computation needed to obtain an image.

VLBI signals are recorded digitally on magnetic tape. The timing of individual samples is controlled by the maser clock and is not affected by mechanical factors such as the accuracy of the tape speed. When the first VLBI system went into operation in the early 1960's, conventional computer tapes were used and the recorded data rate was limited to a few hundred kilobits per second. Today modified home videocassette recorders are employed to record digital samples of the signal at a rate of four megabits per second. This system, called Mark II, provides four hours of uninterrupted recording at a two-megahertz signal bandwidth on a single cassette. More than 25 radio telescopes around the world have recording systems of this type. The newer Mark III recording system developed at the Massachusetts Institute of Technology's Haystack Observatory, largely under the sponsorship of the National Aeronautics and Space Administration, uses a reel-to-reel instrumentation recorder to obtain data rates of up to 224 megabits per second. This recorder provides for signal bandwidths of 112 megahertz, but it consumes a full 10,000-foot reel of tape every six minutes.

The VLBA requires bandwidths of at least 100 megahertz in order to be sensitive to weak celestial radio sources, but this must be balanced against a need to hold down the volume of tape. The Haystack Observatory has devised a new system for the VLBA that has 512 tracks and will be able to handle 256 megabits per second while using up tape more slowly than Mark III. A single 16-inch reel holding eight kilometers of tape will last for more than 12 hours and store about seven trillion bits of data—enough to contain the information found in 1,000 years' worth of issues of a major daily newspaper.

Two reels of tape will normally be recorded every day at each of the antennas and flown to the VLBA Operations Center, where they will be played back and the data combined with data from the other antennas. Up to 20 tapes can be played back simultaneously, so that non-VLBA radio telescopes can also contribute data to enhance the sensitivity and resolution of the VLBA. Even with many baselines the data will contain gaps, which can cause spurious features, called sidelobes, to show up in the computed image. These sidelobes, however, are well understood and can generally be removed from the image by means of a well-tested computer algorithm. More serious errors can be caused by unpredictable effects in the earth's atmosphere, but practical algorithms exist for handling these as well.

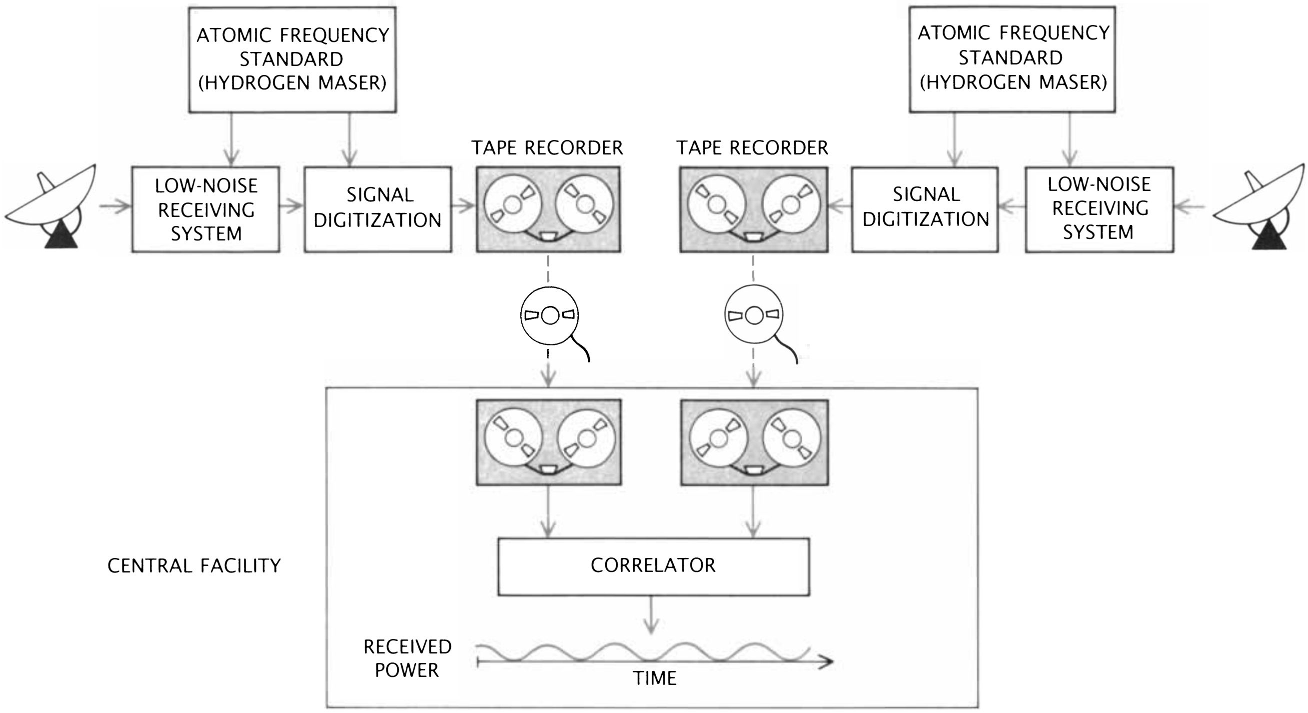
Jets and Stellar Evolution

The VLBA will give high-quality radio images with a resolution of a few ten-thousandths of an arc-second, which is equivalent to the angle subtended by a pea in San Francisco as seen from New York. On an astronomical scale this corresponds to an ability to detect features with diameters as small as 100 million miles anywhere in our galaxy or as small as a few light-years even in the remotest parts of the cosmos. High on the list of targets will be the cores of galaxies and quasars, the most powerful objects known in the universe.

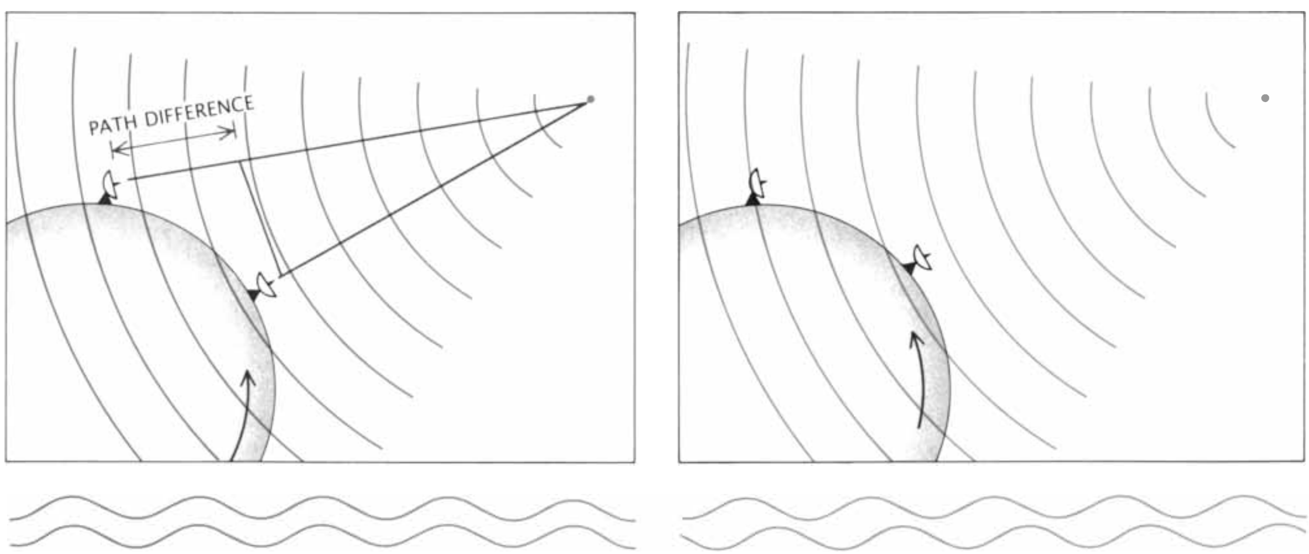
The ad hoc VLBI networks have already enabled radio astronomers to glimpse the inner regions of quasars and active galactic nuclei, where plasma is expelled in narrow jets. These violently energetic events are among the most perplexing problems in astronomy today. The blobs of plasma are ejected at velocities close to the speed of light and create several remarkable relativistic effects. First, the radiation becomes focused into a narrow beam along the direction of motion. If the object is moving close to the line of sight and toward the observer, a relatively weak galactic nucleus may appear to be as bright as a quasar. There is considerable debate among astronomers over whether the intense luminosity of quasars results from this effect.

An even more remarkable consequence of relativistic motion is the illusion that the plasma blobs are moving faster than the speed of light. This arises because the object is moving toward the observer so fast that it nearly keeps up with its own radiation. If the object travels for many hundreds of years, radiation emitted at intervals of hundreds of years will reach the observer only a few years apart. As seen from the earth, the path of travel can be foreshortened and appear to be tens of light-years, giving rise to the illusion that the object has traveled tens of light-years in only a few years. The phenomenon is called superluminal motion and often occurs when a radiating plasma blob is ejected from the nucleus of a quasar. These ejections seem to take place every few years, and it is speculated that the engine driving the activity is a massive black hole. The VLBA's resolution will enable astronomers to understand in greater detail the processes generating the jets.

Closer to home, astronomers are studying objects in our own galaxy in an attempt to understand the life cycle of stars—how they are born and how they die. Here too the VLBA can probe radio phenomena associated with stellar activity that cannot be resolved with conventional radio telescopes. Of particular interest is the intense microwave radiation from cosmic hydroxyl (OH) and water-vapor masers that are found in the gaseous envelopes of very young



VERY-LONG-BASELINE INTERFEROMETRY (VLBI) makes use of widely spaced antennas to observe cosmic radio sources. A hydrogen-maser clock at each antenna synchronizes the observations and provides a frequency standard so that phase relations will be preserved when the data are recorded on magnetic tape. The recordings are taken to a central facility and replayed into the correlator, which combines the signals. The resulting interference pattern yields an image by Fourier transformation.



THE EARTH'S ROTATION causes a pair of radio-interferometer antennas to oscillate between being in phase and out of phase with each other. The amount of rotation needed to change the phase relations depends both on the wavelength of the observed radio waves and on the distance separating the antennas: the shorter the wavelength or the greater the antenna separation, the less rotation required. This in turn determines the amount of detail that the pair of antennas can resolve as they scan the sky.

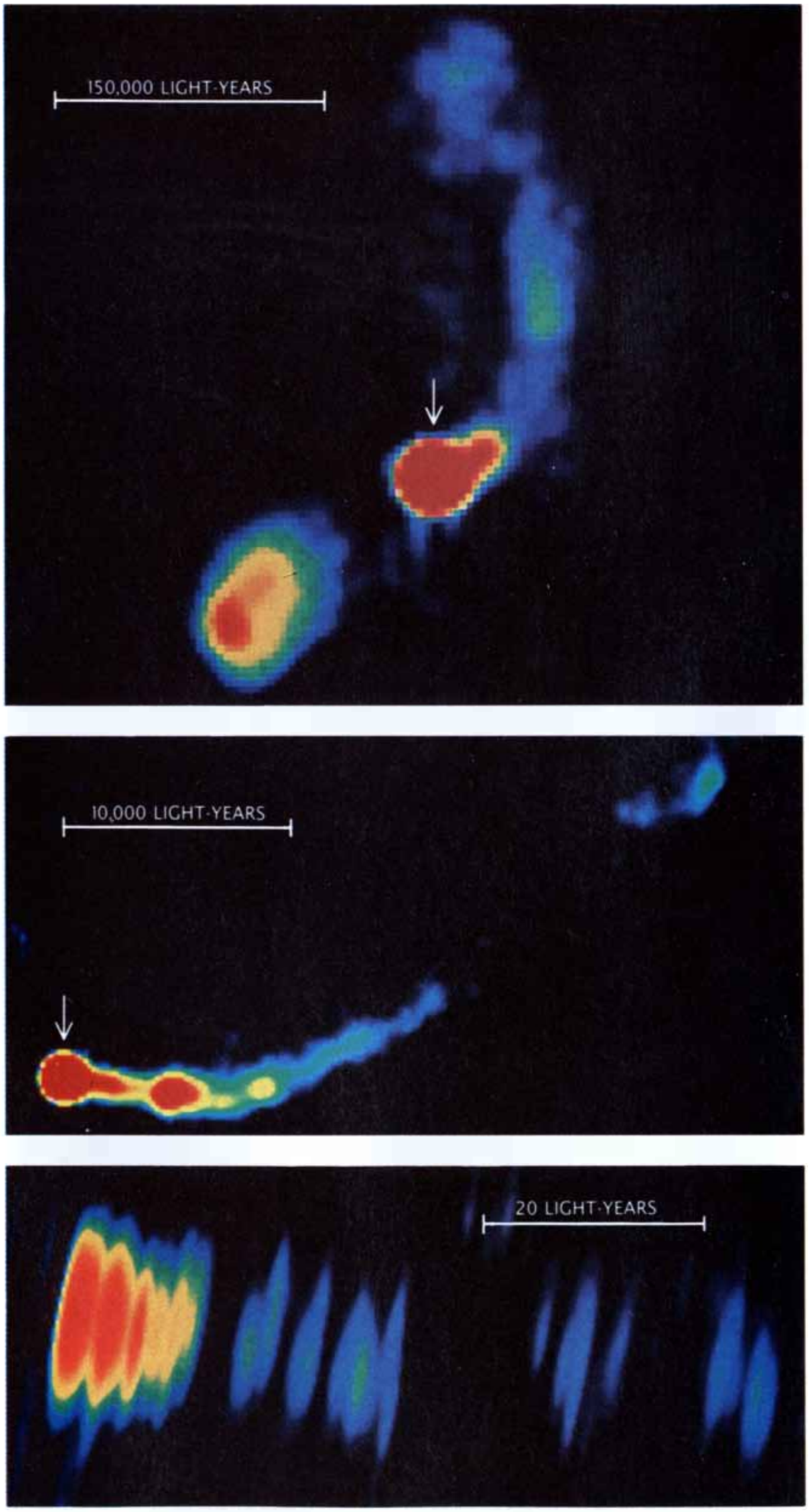
stars as well as in the dusty clouds around aging red giants.

These cosmic masers, among the brightest microwave sources in the sky, emit radiation when clouds of hydroxyl radicals or water molecules are stimulated to drop from a high energy state to a lower one. The excess energy is released as intense narrow-band microwave signals at characteristic wavelengths near 18 centimeters for the hydroxyl radical and 1.3 centimeters for water. Maser sources often contain many separate bright spots whose velocities can be determined by the Doppler shifts of their radiation. The VLBA images of maser motions should reveal much about the dynamics of the turbulent clouds, such as whether they are rotating or are flying apart in the aftermath of a cosmic explosion.

The hydroxyl maser may also act as a sensitive probe of magnetic fields in the maser cloud. Magnetic fields create small differences in the energy levels of the hydroxyl radical. The emitted radiation is split into pairs of spectral lines whose wavelengths are slightly different, a phenomenon called the Zeeman effect. High-resolution observations of the splitting can reveal details of the magnetic field in three dimensions throughout the maser cloud and so make it possible for astronomers to study the role of magnetic fields in causing clouds to collapse and form embryonic stars.

Measuring Cosmic Distances

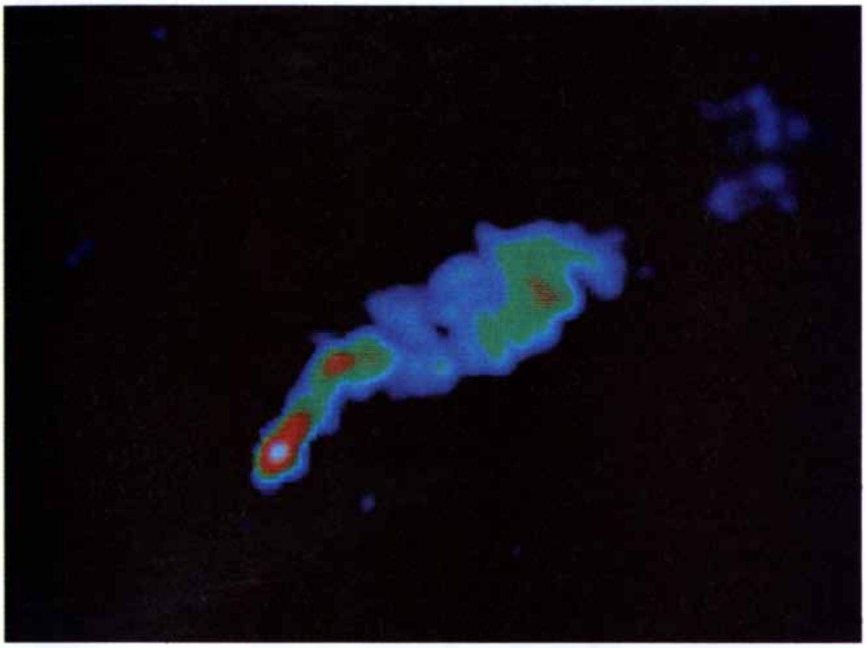
The VLBA will also be able to yield the most accurate measurements yet of cosmic distances, a problem of fundamental importance in astronomy. An accurate scale of the universe is vital to any understanding of its total mass and energy and of its past and future evolution, and yet the overall scale of the universe is uncertain by a factor of two. High-resolution instruments provide a way to determine distance directly. For example, in the case of a supernova remnant that is assumed to be expanding evenly in all directions, the velocity of the expansion can be measured by the Doppler shift in its radiation. Then, by comparing the velocity with observations of the changing diameter, one can infer the distance. The technique was recently applied by an international team of radio astronomers to determine the distance to a supernova that occurred in the Virgo cluster in 1979. VLBI observations gave an angular



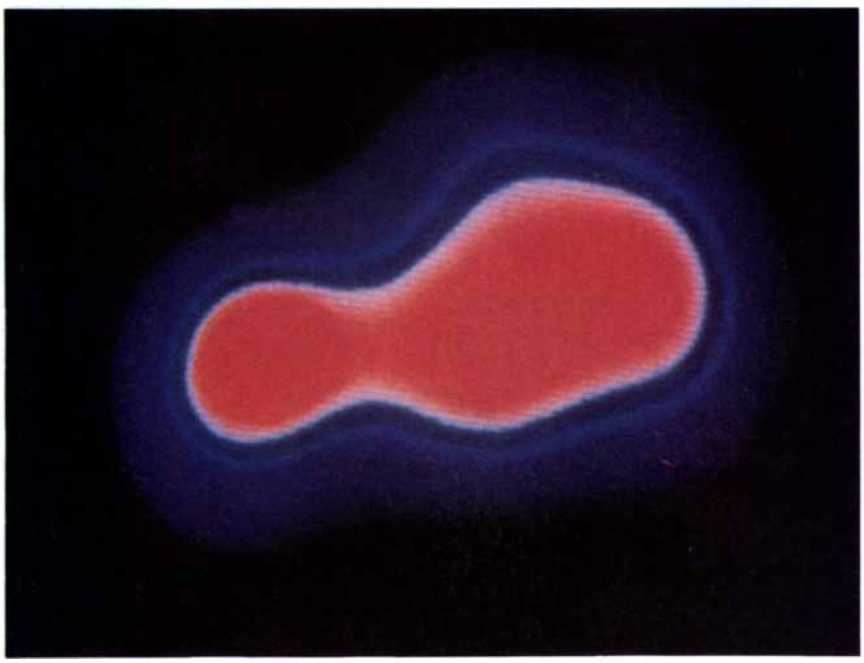
RADIO GALAXY 3C 120, half a billion light-years away, reveals more detail as the resolution of the image improves. The top image, made by the VLA at 1,667 megahertz, has a resolution of nine seconds of arc. In the middle image, made at 15 gigahertz with the array's arms extended to the full 21 kilometers, the VLA resolves features as small as .15 arc-second and shows a narrow jet curving away from the active nucleus. The bottom image, obtained by an 11-antenna VLBI observation made at five gigahertz, reveals details as small as .001 arc-second. The VLBI observations measured the apparent superluminal motion of plasma blobs in the jet as being about seven times the speed of light.

expansion rate of about .003 arc-second per year. Optical and ultraviolet spectra indicated that the supernova is expanding at 11,000 kilometers per second, which corresponds to .003 arc-second per year at a distance of about 60 light-years. So far the technique is about 35 percent accurate, but the VLBA is expected to improve the performance significantly.

VLBI has also provided a highly precise measurement of the relativistic bending of radio waves by the gravitational field of the sun. Originally proposed as one of the three classical tests of general relativity, the effect was confirmed in a landmark experiment in 1919, when stars observed during a total solar eclipse appeared to be deflected by the gravitational field of the sun. The optical observations, however, have been plagued by the difficulties of carrying out experiments in remote corners of the globe during the few minutes when a solar eclipse darkens the sky. Radio waves, on the other hand, can be observed at any time. Using data collected by two VLBI projects, POLARIS (Polar-Motion Analysis by Radio Interferometer Surveying) and IRIS (International Radio Interferometric Surveying), investigators at the U.S. National Geodetic Survey have measured relativistic bending over almost the entire sky. Thousands of observations over several years confirm the accuracy of theoretical predictions to a few tenths of a percent.



POWERFUL ENERGY SOURCE inside quasar 3C 380 appears in this VLBI image as a bright spot less than 10 light-years wide. It is thought the intense energy can be generated only by a black hole. The VLBA will be able to reveal details of such processes.



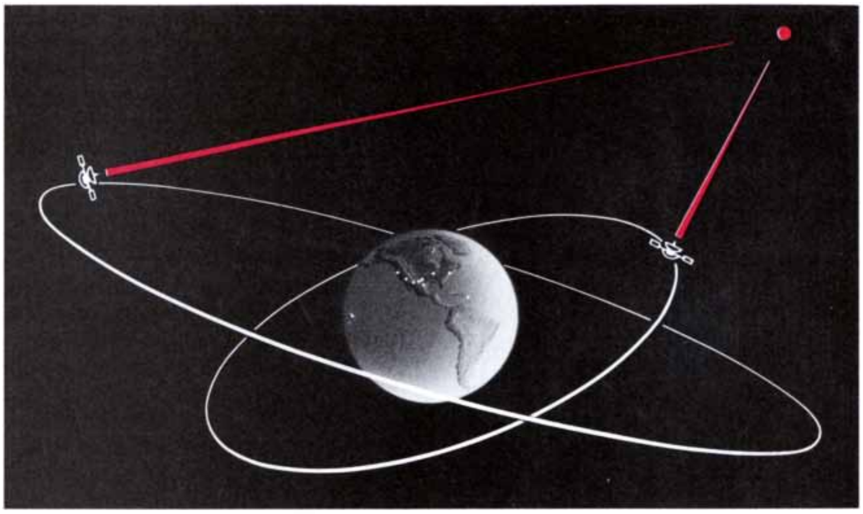
INTENSE RADIO OUTBURST in binary star UX Arietis was recorded in July, 1983. The image, made by a six-antenna VLBI system operating at five gigahertz, is barely able to distinguish the two parts of the binary. The radio emissions are thought to be generated by relativistic electrons accelerated in the stars' magnetic field. The VLBA will be able to observe the phenomenon in greater detail throughout the orbital cycle of the binary.

Measuring the Earth

Although it was conceived as an astronomical tool, the VLBA will be employed to study a variety of important terrestrial phenomena as well, such as plate-tectonic motion, rotation of the earth and distortions in the solid earth caused by tidal forces. These geodetic applications are possible because the response of the radio interferometer depends not only on the source under observation but also on the earth's rotation and the length and position of the baseline separating a pair of antennas. The technique is based on repeated observations of reference sources, such as quasars, so distant that they can be treated as fixed points in the sky. Changes in the earth's shape or rotation alter the results from one observation to the next [see "Studying the Earth by Very-Long-Baseline Interferometry," by William E. Carter and Douglas S. Robertson; SCIENTIFIC AMERICAN, November, 1986].

With geodetic VLBI it is possible to determine the distance between two widely separated antennas with a precision somewhat better than the wavelength at which the observation is made. VLBI can therefore detect changes in the earth's dimensions of less than one centimeter. The Jet Propulsion Laboratory's project ARIES (Astronomical Radio Interferometer Earth Surveying) is making extensive measurements in an effort to relate the incidence of earthquakes to small shifts across fault lines. Not surprisingly, earthquake-plagued countries such as China, Italy and Japan have shown great interest in geodetic VLBI and have already built radio telescopes dedicated to fault-line measurements. Because VLBI fringe oscillations are caused by the rotation of the earth, geodetic observations can measure the earth's rate of rotation in units of time provided by the maser clocks. The length of the day has thus been measured to an accuracy of a tenth of a millisecond, and the location of the earth's spin axis has been pinpointed to within a few tens of centimeters on the earth's surface.

To make the geodetic measurements, one needs to disentangle the terrestrial effects from the astronomical data. This is a complex problem and is made more difficult by the fluctuations in the radio sources, the unpredictable effects of the earth's atmosphere and ionosphere on the radio signals, the slow drift of the atomic clocks at the various stations and even the relativistic bending of radio waves by the sun. In order to refine the important geodetic and geophysical parameters, radio telescopes throughout the world have combined forces under the aegis of NASA's Crustal Dynamics Project and the POLARIS and IRIS projects mentioned above. With the VLBA it will be possible to obtain accurate images of the cosmic radio sources that provide the geodetic reference points. At the same time the VLBA will provide regular observations to supplement and extend the geodetic VLBI networks already in operation.



ORBITING RADIO TELESCOPES will extend interferometer baselines into space. During the 1990's the U.S. and Europe plan to launch *Quasat* into orbit three earth-diameters away, and the Soviet Union plans an even higher orbit for its *Radioastron* satellite.

VLBI in Space

Still another application of the versatile VLBA will be in interplanetary navigation. VLBI networks demonstrated this capability with spectacular success during the recent Soviet Vega mission to Venus and Halley's comet. In the largest international VLBI venture yet organized, a network of 20 antennas around the world tracked the two spacecraft deployed in the mission. Each of the spacecraft, which reached Venus in June of 1985, released a balloon carrying a 1.7-gigahertz transmitter into the planet's forbidding atmosphere. The VLBI network tracked the balloons as they were buffeted by the Venusian winds and measured gales of up to 140 miles per hour.

Following the encounter with Venus, the two spacecraft continued on to rendezvous with Halley's comet on March 6 and 9, 1986, a week before the scheduled flyby of the European *Giotto* spacecraft. European Space Agency scientists hoped to steer *Giotto* to the sunny side of the comet in order to take pictures, but because they could not be sure of the comet's precise location, there was a real danger that their cameras would end up on the dark side. Optical images of the comet made by Vega, combined with VLBI data on the spacecraft positions, enabled the Europeans to make last-minute corrections to *Giotto*'s trajectory and bring it to within a few hundred miles of the sunlit side of the icy core. There *Giotto*'s cameras were able to make spectacular photographs of Halley's comet on March 14, during its closest approach, just before communications with the spacecraft were cut off by collisions with the comet's debris.

Even when the VLBA is used in conjunction with other radio telescopes around the world, its resolution will ultimately be limited by the size of the earth. To achieve higher resolution the baselines will have to be extended into space, perhaps to the moon or even to the planets. Space-based VLBI will present a challenge to engineers, who must develop large, precise antennas and sensitive radio receivers that can operate unattended in the harsh space environment. The feasibility of space VLBI has already been demonstrated by a team of American, Australian and Japanese scientists, who employed a small antenna aboard a NASA satellite as one element of an earth-space VLBI system.

Plans for dedicated VLBI satellites are already being discussed in the U.S., Western Europe, Japan and the Soviet Union. It has been proposed that NASA and the European Space Agency jointly launch a satellite to be called *Quasat*, which would carry a 10-to-15-meter antenna into earth orbit by the middle of the next decade. Recent setbacks to the NASA space-science program have made U.S. participation in this project uncertain. The Soviet Union has also announced a space VLBI program and expects to launch two or three *Radioastron* satellites in orbits of up to 75,000 kilometers. European and American scientists have been invited to take part, but the U.S. Government's reluctance to share advanced space technology with the Soviet Union may limit U.S. involvement.

In the more distant future, some Soviet scientists envision building giant antennas, each as much as several kilometers across, to be placed in orbit around the sun. This array would provide baselines up to several hundred million miles long and an angular resolution that could be better than a millionth of an arc-second. A radio telescope of such power would open up a new frontier of astronomy. In theory it would enable scientists to observe "sunspots" on other stars in our galaxy and to resolve features in neighboring galaxies comparable to the supposed size of black holes. Fascinating as it is to speculate about the future, there remain many uncertainties about the feasibility and effectiveness of such an array. These issues should become clearer as we gain experience with the VLBA and earth-orbiting antennas.

Intertidal Fishes

Fishes that live between the tides are alternately buffeted by waves and isolated in pools and on mud flats. Anatomy, physiology and behavior suit them to their rigorous habitat

by Michael H. Horn and Robin N. Gibson

The intertidal zone—the band of shoreline that is between the low-tide mark and the high-tide mark—is a demanding environment. Roughly twice a day, at low tide, the intertidal zone is cut off from the open ocean. The water that remains is left in isolated tide pools or under rocks, or else it has combined with the substratum to form mud flats. When the water returns at high tide, the intertidal zone is submerged to rejoin the broader ecosystem of the ocean. Any animal that lives in the intertidal zone must be able to spend much of its time either completely out of the water or at least partly exposed to the air. The periods during which it can feed are set by the tidal cycle, and it must withstand wide variations in water chemistry and the nearly constant wave action and turbulence that are characteristic of the zone.

The fixed and slow-moving animals inhabiting the intertidal zone, such as barnacles, limpets and periwinkles, have been studied intensively, primarily because they are abundant and accessible. Intertidal fishes have been more difficult to study, in part because of their mobility and in part because they are generally well hidden. Many of them are camouflaged by protective coloration, and at low tide many species live under rocks or within clumps of vegetation (although some species, such as the tropical mudskippers, can live in the open on mud flats).

Nevertheless, in the past 15 to 20 years our knowledge of these fishes has increased to the point where it is now possible to give a fairly coherent picture of how they make a living in this unusual habitat. Our work and studies by other investigators have revealed that intertidal fishes are remarkably well suited to the environment in which they live; they are not simply ocean animals that have been stranded by the ebbing tide. Indeed, these fishes have evolved to become integral components of the intertidal ecosystem, with each species of fish occupying a characteristic vertical distribution and type of habitat on the shore.

Many intertidal fishes display the specialized anatomical features characteristic of fishes dwelling on the bottom of shallow, turbulent waters. The most obvious of these features is small size. Intertidal fishes are rarely longer than 30 centimeters, and most of them are less than 20 centimeters long. Their small size enables them to occupy holes, crevices and spaces under rocks, and it reduces the risk of their being swept away by surge or waves. The ability of many intertidal fishes to live in confined spaces is also enhanced by such characteristics as very thin bodies (in the gunnels and pricklebacks) or flattened, horizontal profiles (in the sculpins and clingfishes).

In addition to specialized body shapes and sizes, many intertidal fishes have distinctive fin structures. For example, the highly terrestrial mudskippers can raise themselves on their paired pectoral fins when they are moving about on the land. In some intertidal fishes (such as clingfishes, gobies and snailfishes) the pelvic fins are modified to form suction cups by which the fishes can attach themselves firmly to the substratum. On the other hand, in the fishes that are most highly modified for dwelling in crevices or holes (such as certain gunnels and pricklebacks) the pectoral and pelvic fins are greatly reduced and the dorsal and anal fins are long, low and often united with the tail fin.

The skin of intertidal fishes is generally tough, and so it can withstand repeated scraping against the substratum. Some of the fishes (such as blennies and clingfishes) have no scales; in others (such as gunnels and pricklebacks) the scales are greatly reduced and in still others (the gobies) they are attached very firmly. Many of the fishes secrete large amounts of mucus, which may provide lubrication in confined spaces. (It may also help to reduce the loss of water when the fishes are exposed to the air.)

A standard feature of intertidal fishes is cryptic coloration. Flatfishes on sandy shores (plaice and the California halibut, for example) are renowned for their remarkable ability to match their pigmentation with the colors and patterns of the bottom. On rocky shores, where the background colors can be more varied, intertidal fishes exhibit a broad range of colors and patterns. Species that live within algal vegetation usually have striking colors that match the surrounding seaweed. For instance, the gunnels *Apodichthys flavidus* and *Xererpes fucorum* can range in color from tan to bright green to deep red, depending on the color of the algae in which they live. These fishes, which change color slowly, apparently gain pigments from their invertebrate prey, which also inhabit seaweeds. Other fishes can change color within seconds to blend in with sessile invertebrates or encrusted rock surfaces.

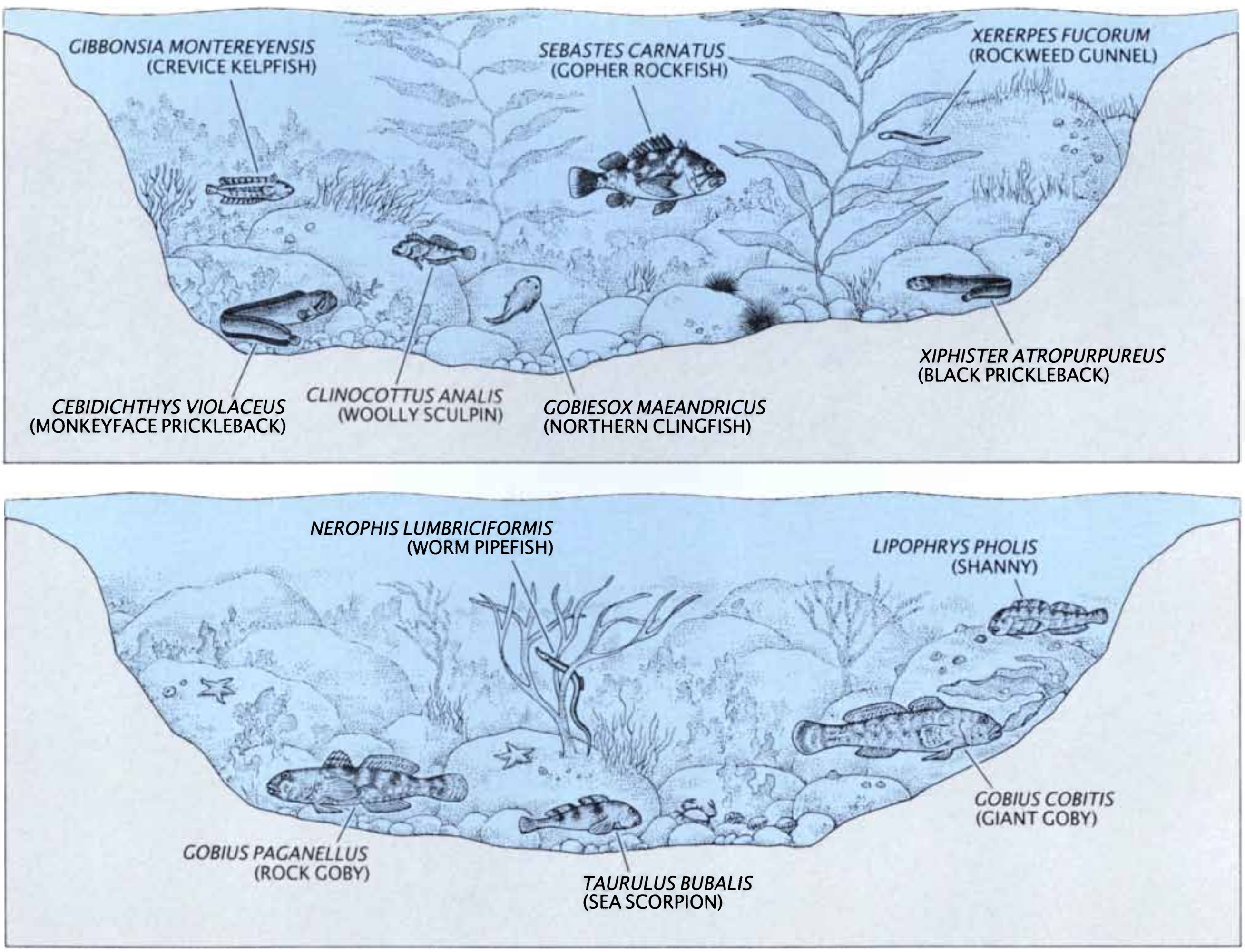
Many intertidal fishes have negative buoyancy (they are heavier than seawater), which enables them to lie effortlessly on the bottom, where cover is nearby and water velocities are lowest. Such fishes have either no swim bladder (the gas-filled organ characteristic of most bony fishes) or a greatly reduced one. The relatively high specific gravity of intertidal fishes helps to explain their labored swimming, which usually consists of brief excursions from cover or short darts from one place to another.

The intertidal region experiences frequent and dramatic fluctuations in environmental conditions, threatening its inhabitants with desiccation during exposure to air and confronting them with variations in oxygen availability, salinity and temperature that would be lethal to other species. Resident fishes must be behaviorally and physiologically equipped to cope with such great changes. Even intertidal species vary in their tolerance to these stresses, particularly to desiccation. The variations help to determine the vertical distribution of species on the shore.

Several studies have shown that many intertidal fishes can tolerate considerable loss of water. For example, William H. Eger, while a graduate student at the University of Arizona, found that certain clingfishes from the Gulf of California can survive for as long as 93 hours out of the water if the humidity is high (90 percent) and can sustain water losses as great as 60 percent of their total water content in environments where the humidity is low (5 percent). This is a higher loss of water than can be tolerated by most of the amphibians that have been studied so far. (In fact, certain intertidal fishes, the tropical mudskippers, virtually *are* amphibians: unlike many other intertidal species, they are often active out of water and may spend between 80 and 90 percent of their time on land.)

The rate of desiccation may be slowed in some species by such anatomical features as a thickened epidermis and the presence of mucus-secreting cells in the skin, but intertidal fishes apparently have no other physiological mechanisms for reducing or preventing water loss. Indeed, live and dead fish exposed to the air lose weight at about the same rate. It seems, then, that behavior—simply trying to avoid desiccating conditions—plays an important role in survival. Most species remain inactive when they are exposed at low tide. Fishes that are active out of water, such as the mudskippers and rockskippers (which are not related to each other), ensure against excessive drying of the skin and respiratory surfaces by returning frequently to the water or remaining within reach of waves or spray.

Changes in the supply of oxygen cause two kinds of problem for intertidal fishes. First of all, the fishes must somehow obtain oxygen even when they are out of the water for several hours. The availability of ox-



INTERTIDAL FISHES common along the California coast (*top*) and the Atlantic coast of France (*bottom*) are shown in idealized tide pools at high tide, when they are active. At low tide, when the water level in the tide pools falls or they dry out entirely, the fishes are usually hidden in seaweed and under rocks. Although each site is host to a distinctive set of species, the fishes share anatomical features that suit them for hiding in tight spaces and withstanding wave action, among them small size, thin bodies or flattened profiles, and pelvic fins (the paired fins on their underside) reduced in size or modified for clinging to the bottom.

ygen is not the problem, of course, because the concentration of oxygen is much higher in air than in saturated water. The difficulty comes in acquiring the atmospheric oxygen. In general, gills tend to collapse in the air; their thin, flexible structure is much better suited to operating underwater. The problem is solved in mudskippers and rockskippers by shorter, thicker gill filaments, which prevent collapse of the gills and enable them to function for breathing air. Most species that can breathe air have other specialized features, such as increased numbers of blood vessels in the skin and in the lining of the mouth and pharynx.

In these air-breathing species the rate of respiration in air is nearly equal to the respiration rate in water. Furthermore, according to Christopher R. Bridges of the University of Düsseldorf, the concentrations of lactate in the blood and muscle of intertidal species show that the rate of anaerobic metabolism (metabolism that does not directly involve oxygen) does not increase when the fishes are exposed to the air, indicating that the supply of oxygen to the fishes' cells is not greatly reduced when they must breathe air. The fishes can therefore function out of water without any major changes in their mechanisms of metabolism.

The second kind of respiratory problem confronts intertidal fishes that inhabit isolated tide pools high above the low-tide mark. Photosynthesis by seaweeds in the pools and respiration by the resident plants and animals cause strong diurnal fluctuations in the amounts of dissolved oxygen and carbon dioxide

ROCKY COAST, with its tide pools, boulders and seaweeds, offers a diversity of habitats for intertidal fishes. It is also a turbulent and dynamic environment: the fishes that make their home there must cope with waves, surge and the physiological stresses imposed by the ebb and flow of the tide. The photograph was made on Washington's Olympic Peninsula by Anne Wertheim.

in the pool water. During the day seaweed photosynthesis can supply amounts of oxygen exceeding the respiratory demands of a pool's inhabitants, resulting in high levels of oxygen and low levels of carbon dioxide. At night, when photosynthesis ceases, respiration by the pool's plants and animals depletes the oxygen supply and increases the level of carbon dioxide.

During the day tide-pool fishes do not seem to change their rate of oxygen consumption in response to prevailing conditions. At night, however, some species do reduce their oxygen consumption to make up for the poor availability of oxygen. When oxygen reaches critically low levels, some species have been observed to ventilate their gills with the thin film of water at the surface of the pool, where there is more dissolved oxygen; in the laboratory, fishes in water that has an extremely low level of dissolved oxygen sometimes emerge from the water and climb onto exposed surfaces.

Along with unusual anatomical and physiological traits, fishes in the intertidal zone have distinctive behavior patterns. For example, the reproductive process of most intertidal fishes follows a generalized pattern. First the male selects a spawning site in a sheltered location (under a stone or within vegetation), which in many species is within a territory the male defends before and after spawning. Then he attracts one or more females to the site; each female deposits a single batch of eggs on the substratum. The male fertilizes the eggs and then usually tends them until they hatch.

The male's need to choose and protect a cryptic spawning site puts a number of constraints on his courtship behavior. He cannot leave the site to search for mates, and so he must attract females to him. In order to attract females the male relies on numerous displays. For example, some male blennies advertise themselves by performing vertical loops, and they induce the female to enter the spawning site by vigorous head movements. The ability of the male to gain attention by displays is augmented by tentacles, crests and distinctive color patterns. Certain species also utilize pheromones, or specific olfactory signals, for communication between the sexes.

Parental care of the eggs is the rule, whether the eggs are attached to the substratum (as they are in most species) or are formed by the parent into

CRYPTIC COLORATION conceals a sculpin, a species of *Clinocottus* (*top*), and a gunnel, *Xererpes fucorum* (*bottom*), against their backgrounds, respectively a sponge and a clump of seaweed. Such camouflage, which probably affords protection against predators, is common among tide-pool species. The sculpin can change color to match its surroundings by means of specialized pigment cells; the gunnel apparently takes its green color from invertebrates it preys on. The sculpin was photographed in its habitat by Anne Wertheim, and the gunnel—a captive specimen—by one of the authors (Horn).

a cohesive but unattached mass. Usually it is the male who guards the eggs, although in some species it is the female and in others it is both parents. Parental attention safeguards the eggs from predators and probably keeps them from being swept away by surge and wave action.

After hatching, the larvae of most intertidal fishes develop in the ocean itself, as members of the plankton. How they return to the intertidal zone has remained largely a mystery. One possible answer involves the slicks (long strips of smooth water running roughly parallel to the shore) that are produced by the action of internal, or subsurface, waves in the ocean. Recent research has shown that slicks contain higher concentrations of fish larvae than adjacent rippled areas do. Slicks are common coastal features in regions where the temperature or the salinity of the water column changes sharply with depth, and they generally move slowly toward the shore. Perhaps they provide a mechanism by which the larvae of some intertidal fishes can return to the intertidal zone.

On the other hand, work by Jeffrey B. Marliave of the Vancouver Public Aquarium suggests that the larvae of some species may not need to be brought back from more distant waters. He found that larvae of some species living in rocky areas remain close to the shore, resisting any forces that might disperse them offshore or even along it. In these species individuals from different regions would have little opportunity to mingle, and so even populations that are relatively close to one another on the coastline might be genetically isolated.

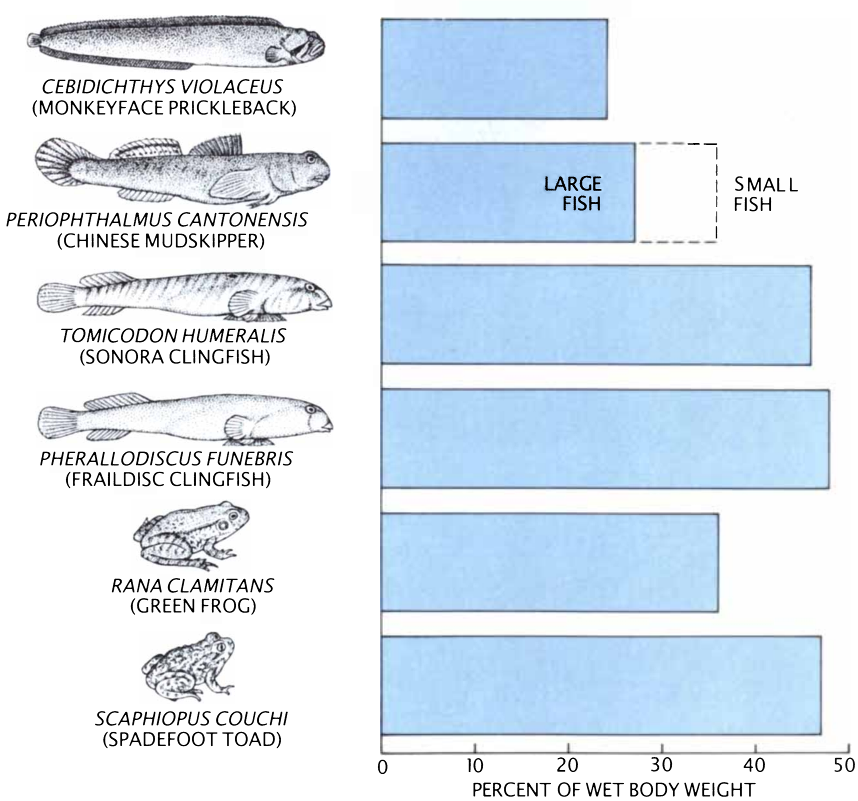
Another common feature of many intertidal fishes is their homing ability. Intertidal fishes must undertake periodic foraging excursions, and as a result they risk being stranded in unfavorable locations at low tide. Many species have developed homing abilities that enable them to return to a particular location, generally a tide pool, that provides suitable refuge. (Species that do not need to find tide pools but instead live mainly under rocks or in vegetation generally do not have such homing ability.) The ability to home and to become familiar with the topography of a particular area may also help intertidal fishes to escape predators and turbulent weather at high tide.

Most studies of homing behavior done so far have involved the tidepool sculpin *Oligocottus maculosus*. A standard technique is to label fish captured in a given area by attaching colored plastic beads to them and then move the marked fish to a different location. The investigator later revisits the site from which the fish were originally taken to see how many can be recaptured there. John M. Green of Memorial University of Newfoundland and Gwenneth J. S. Craik, then at the University of British Columbia, found that as many as 80 percent of *O. maculosus* return to their native pool after having been displaced by distances as great as 100 meters. Young fish had poorer homing ability than older ones and were unable to retain their homing ability for as long. Some older fish could still return home after having been kept in the laboratory for as long as six months before being released.

Both vision and olfaction appear to be important for homing in the tidepool sculpin; it remains to be determined whether one sense is more important than the other. Craik has concluded that either sense can be sufficient for older fish, whereas both of them are necessary for younger specimens.

Definitive studies on the details and significance of homing have yet to be done, but new techniques based on ultrasonic telemetry are finally making such research possible. In a recently completed study telemetry enabled one of us (Horn), working with Scott L. Ralston of Deep Ocean Work Systems in San Pedro, Calif., to track the movements of several monkeyface pricklebacks (*Cebidichthys violaceus*) continuously for periods of up to two weeks. A small transmitter attached to the inside surface of each fish's gill cover emitted high-frequency sounds, which were detected by three hydrophones deployed on the sea floor. The hydrophones relayed the signals to an onshore receiver and microcomputer, which determined the fish locations, displayed them on a screen and stored the information for later analysis.

We found no evidence of homing; our results suggest instead that the monkeyface prickleback stays with-



TOLERANCE TO DESICCATION of some intertidal fishes is comparable to that of amphibians. The bar graph shows the percent of normal weight to which each species could be reduced without killing the animal. The ability to survive water loss is one factor determining how high on the shore a species is found and its level of activity at low tide. Mudskippers, for example, are able to crawl on exposed mud flats and breathe air.

in a home range of no more than a few square meters of intertidal habitat. The species also proved to be sluggish: each individual we tracked was active for a total of only about five minutes a day, mostly during the incoming tide. We look forward to applying the technique to other species. Increasing miniaturization of transmitter components should soon make it possible to track fishes even smaller than the pricklebacks we followed, which ranged from 23 to 41 centimeters in length.

It ought to be advantageous for intertidal fishes to sense the state and motion of the tide, in order to synchronize their activity with its regular, twice-daily rhythm. Conceivably the fishes could do so by responding directly to such signals as changes in temperature, salinity, turbulence or light intensity. Yet there is another mechanism: recent studies by one of us (Gibson) have shown that even when such potential indicators are held constant in the laboratory, some fishes, from a variety of families, still display rhythms of behavior that are related to the tidal period. They alternate periods of rest at the expected time of local low tide with bouts of activity at or near the expected time of high tide.

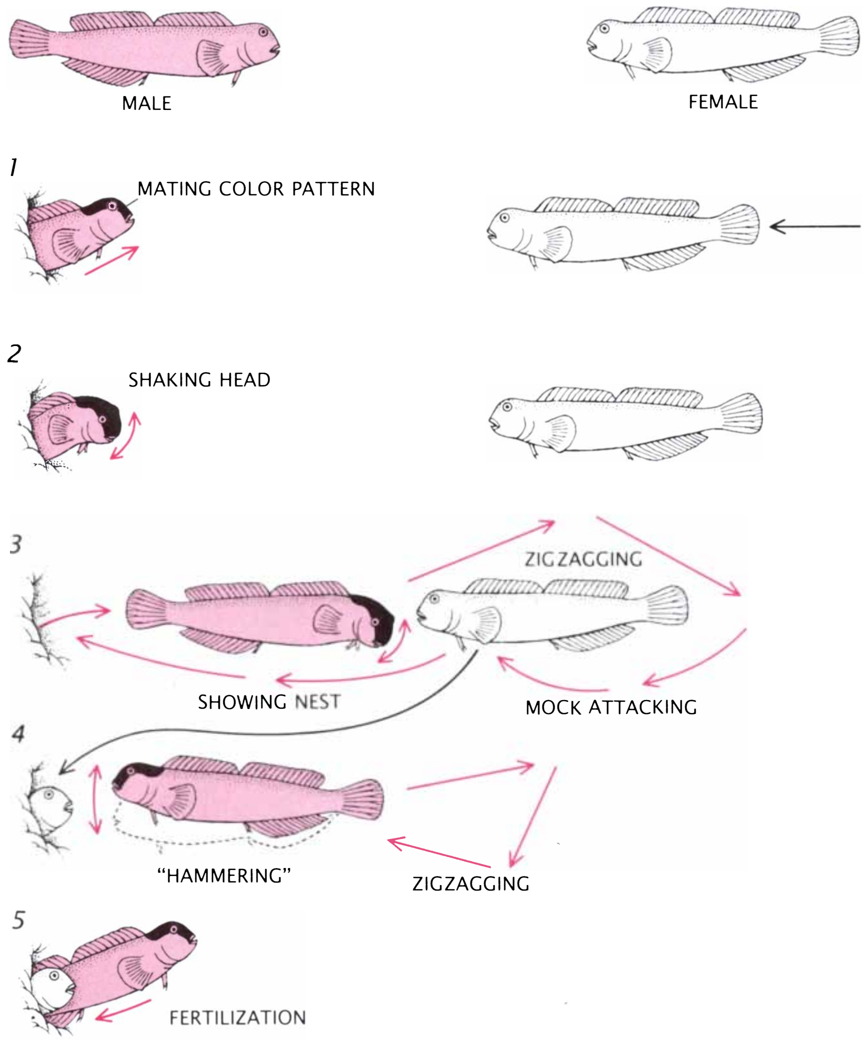
We do not know exactly how such an endogenous rhythm comes to be established, although laboratory investigations show that the changes in hydrostatic pressure associated with the rise and fall of the tide are at least partly responsible. Many intertidal fishes have no swim bladder, the organ usually considered to be responsible for the detection of pressure by bony fishes, and so how they detect these slow pressure changes remains a puzzle.

In nature the fishes' responses to the tides are sometimes modulated by daily cycles of light and dark, and perhaps also by an inherent circadian (approximately 24-hour) rhythm. Field studies have demonstrated that many species are strongly diurnal in their patterns of activity. This is probably related to their dependence on vision for such major activities as foraging, mating and avoiding predators. Knowledge of daily rhythms in intertidal fishes, like knowledge of their homing behavior, will be greatly increased by studies that exploit ultrasonic telemetry.

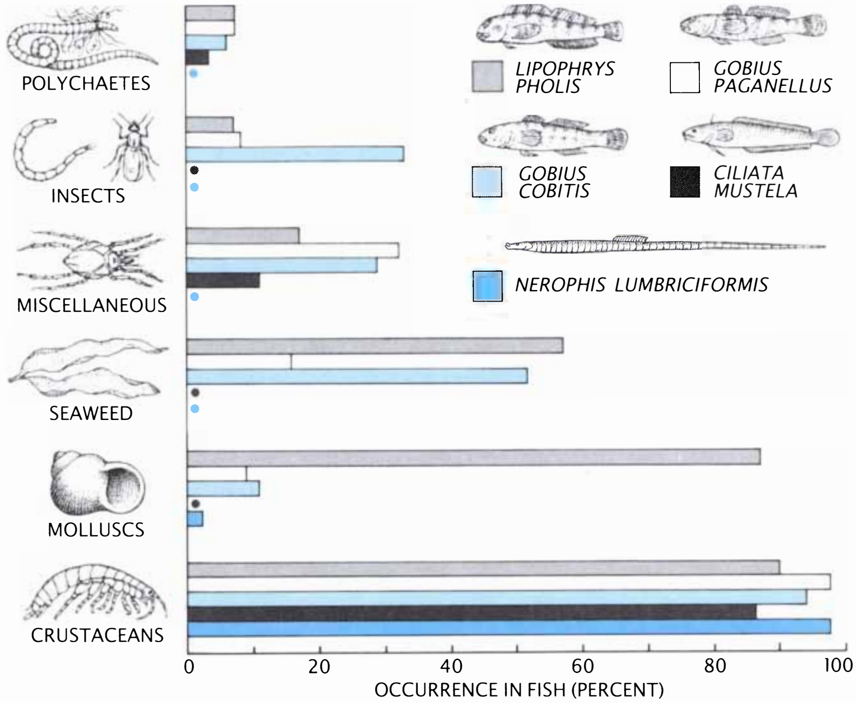
In addition to its other characteristics, the intertidal zone presents a particularly rich area for feeding. The great mass and diversity of living material in the intertidal zone, particularly on rocky shores in temperate waters, is matched by the variety of feeding styles found there. The complex interrelations of the intertidal food chain are further complicated because many species change their eating habits seasonally or annually and still others change their tastes as they age.

Most intertidal fish species are carnivorous, but some eat a combination of plant and animal foods and a few are strictly herbivorous. The intertidal fishes that live in temperate climates and eat primarily seaweeds are particularly interesting. How do they extract an adequate diet from a food source that is so relatively low in nutrients (particularly protein) and whose cell walls consist of cellulose, a largely indigestible material? Indeed, some marine biologists have suggested that these fishes may not digest seaweed tissue at all but rather may feed on small organisms on the seaweed's surface.

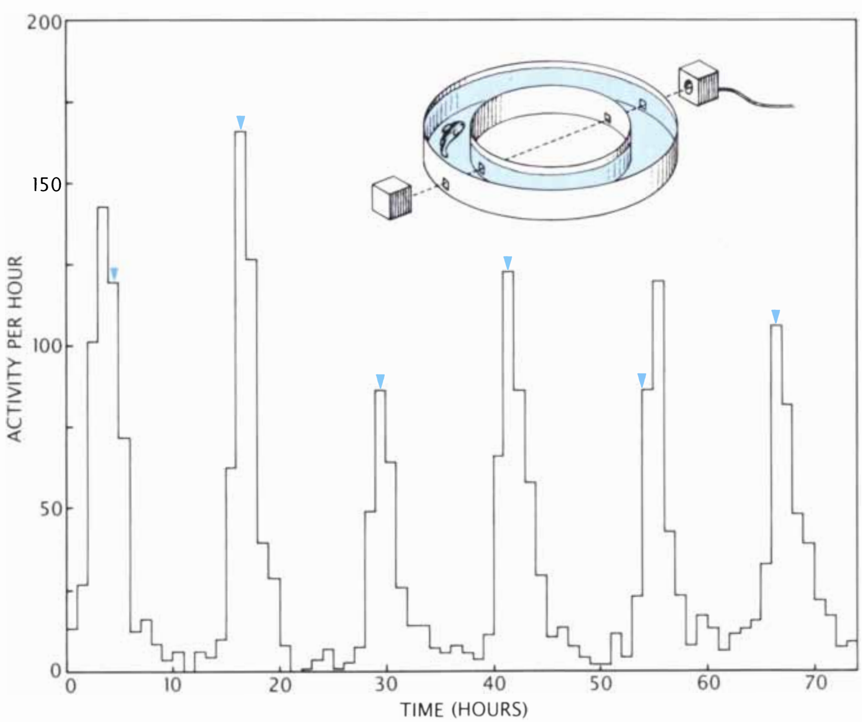
Some insight has been gained into the feeding habits of these fishes from observations of the monkeyface prickleback and another California prickleback, *Xiphister mucosus,* by one of us (Horn) with Steven N. Murray and Margaret A. Neighbors of the California State University at Fullerton. These fishes begin life as carnivores, but when they are less than a year old, they become complete herbivores, feeding on certain species of red and green seaweeds, particularly those that are relatively rich in protein or carbohydrate. While he



COURTSHIP RITUAL of the blenny *Lipophrys canevae* enables the male to attract a female to his concealed nest, where the eggs will be relatively secure against wave action and predators. The male emerges (*1*), catches the female's attention by shaking his head (*2*) and swims out to court her and coax her to the nest (*3*). She lays eggs (*4*); he joins her to fertilize them (*5*). Sheltered nests and elaborate courtships are common in intertidal fishes; this ritual was described by E. F. Abel of the University of Vienna.



DIET OF INTERTIDAL FISHES is shaped by the diversity of invertebrates and plants on rocky shores. The bar graph shows the frequency with which various foods were found in the stomach of intertidal species at Roscoff, France. Most of them consume a variety of invertebrates, small crustaceans in particular, and a few also browse on seaweeds.



INTERNAL ACTIVITY RHYTHM of the shanny (*Lipophrys pholis*), measured in the laboratory, shows regular peaks matching the expected times of high tide (*triangles*) in the fish's natural environment. The rhythm—characteristic of many intertidal species—was recorded under constant conditions in an apparatus that confined the fish in a circular trough (*top*); a photocell registered an activity "count" each time the fish interrupted a beam of infrared light. The internal clock governing such activity rhythms may be set by changes in water pressure that accompany the rise and fall of the tide.

was a graduate student at Fullerton, Kevan A. F. Urquhart demonstrated that the monkeyface prickleback digests the seaweed not by breaking down its cell walls with specialized enzymes but by retaining the food in its gut for a long time (about 50 hours) and gradually leaching carbohydrate and protein from the seaweed with acidic stomach fluids.

Little is known about the impact of intertidal fishes as predators on the plant and animal populations of rocky shores. Limited evidence suggests that in such places as the Great Barrier Reef fishes that browse in the intertidal zone are the major factor controlling the abundance and composition of the intertidal biota. Investigation of the rates at which intertidal fishes of California's rocky coast consume various kinds of prey has led Gary D. Grossman of the University of Georgia to propose that these fishes exert strong selective pressures on many species of intertidal algae and invertebrates. Progress in this area of investigation has been slow because of the difficulty of manipulating (or even determining) the numbers and distributions of intertidal fishes.

The evolutionary forces underlying the colonization of the intertidal zone by fishes are far from being completely understood. These species are able to cope adequately with the demands of their habitat: they tolerate its dehydrating conditions, withstand its characteristically extreme changes in water chemistry, breathe air when necessary, feed on a wide variety of food resources and guard their eggs against both physical and biological dangers.

But if the invasion of the intertidal zone by fishes is to be understood in the context of natural selection, one must ask what the advantages are of living there. It has not been possible to test whether the intertidal zone offers lower levels of competition and predation than exist in the adjoining subtidal zone. It is often suggested that the main advantage of living in the intertidal zone is the opportunity to escape from aquatic predators, but at low tide, when the danger from aquatic predators is least, intertidal fishes may become vulnerable to terrestrial and aerial predators. The relative advantages of the intertidal zone and the impacts of different kinds of predators are difficult but significant areas of study that can be explored only by means of carefully conducted field experiments.

The Not-So-Rare Earths

The rare-earth elements, on which electronic, metallurgical and glass industries depend, are not all that scarce in minerals. The elemental abundances reveal the geochemistry that leads to a mineral's formation

by Gunter K. Muecke and Peter Möller

Their generic name notwithstanding, the rare earths are neither rare nor earths (metal oxides). They are metallic elements, and all but one are more abundant in the earth than gold, silver, mercury or tungsten. (They were once considered to be rare earths, because they were first isolated in oxide form from uncommon minerals.) The rare-earth elements are in fact ubiquitous, being present in low concentration in virtually all minerals.

Yet extracting the rare earths from common minerals would be prohibitive. It is therefore fortunate that a handful of less common minerals do exist in which the rare-earth concentration is high enough to allow the economical extraction of the elements, because the elements are key ingredients in the manufacture of numerous modern products. Cerium and erbium are components of high-performance metal alloys. Neodymium, holmium and dysprosium are needed for certain types of laser crystals. Samarium is a key component of the strongest known permanent magnets, which have made new electric-motor designs possible. The magnetic properties of ytterbium and terbium are applied in magnetic-bubble and magneto-optic devices, which store computer data. Europium activates the red phosphor of color-television screens. Perhaps the most intriguing application of the rare earths is in the manufacture of the ceramics that become superconducting at remarkably high temperatures. Lanthanum was a component of the first such ceramic; investigators have since reported that replacing lanthanum with gadolinium results in superconductivity at even higher temperatures. Rare-earth elements also catalyze the refining of petroleum and regulate the refractive index of glass lenses and optic fibers. (There are also many humble, low-technology products made with rare earths. They are alloyed with iron, for example, to produce the material for cigarette-lighter flints.)

The fact that the rare-earth elements are not uniformly dispersed on the surface of the earth is a boon for mining and ore-processing companies as well as for geochemists. The relative abundances of the rare-earth elements in minerals reflect the interaction of the unique chemical properties of the elements with diverse processes that take place in the upper mantle and crust of the earth. Investigators studying the pattern of rare-earth distribution in minerals can thereby gain geologic knowledge that benefits earth science and ultimately modern industry.

Not only is the term "rare earths" a misnomer but also there is some disagreement as to what elements exactly it encompasses. Many chemists think the term subsumes scandium, yttrium and the lanthanides: the elements that have atomic numbers 57 through 71 [*see illustration on page 74*]. Our discussion centers specifically on the lanthanides, although for convenience we shall continue to use the terms rare earths and rare-earth elements interchangeably. (We also in effect exclude promethium, since it has no stable isotopes and is not found in nature.)

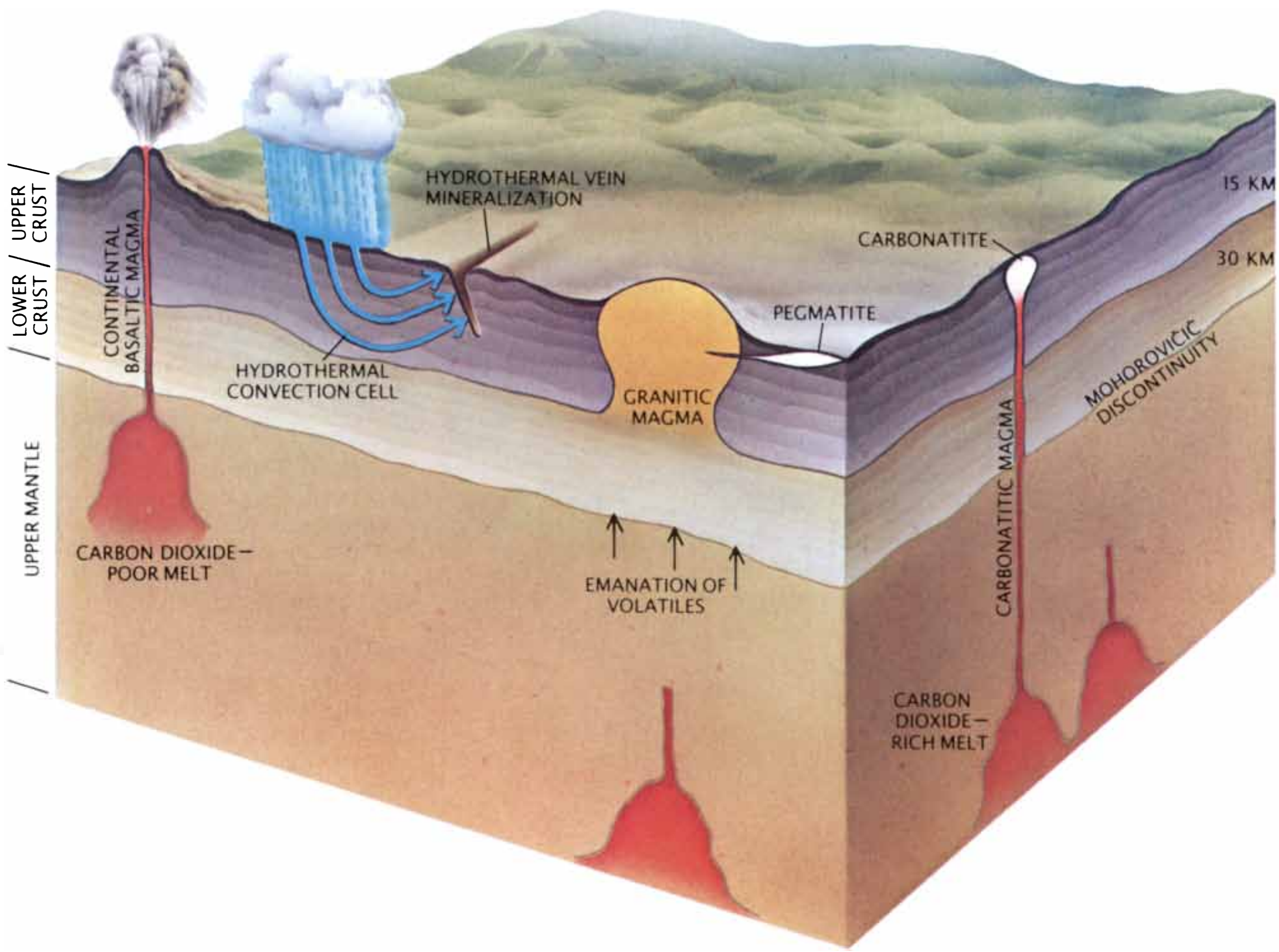
The reason for the terminological confusion arises in part from the elements' reactiveness, which makes it difficult to isolate them in pure form, and the similarity of their chemical properties, which makes it difficult to separate one from another. Indeed, although a description of an unusual mineral containing rare earths was first published in 1788, it was not until 50 years later that the Swedish surgeon, chemist and mineralogist Carl Gustav Mosander demonstrated that the earths isolated from the minerals were in fact not elements at all but mixtures of elemental oxides. The subsequent isolation and characterization of all the individual rare earths in elemental form entailed a long and tortuous process that was finally capped in 1947, when promethium was separated from the fission products of uranium.

The rare earths' chemical coherence, or similarity, is a consequence of the peculiarities of their atomic structure. An atom can be envisioned as a positively charged nucleus surrounded by several electron "shells" (labeled by number) that are built up from "subshells" (labeled by letter) corresponding to the various allowed quantum states of the atom's electrons. Valence, or bonding, electrons are usually in the atom's outermost electron shell, and they determine the chemical properties of an atom. Normally atoms of neighboring elements in the periodic table differ not only in nuclear charge but also in the type of valence electrons each atom has. Atoms of all the rare-earth elements, however, keep the same number and type of valence electrons in their outermost, sixth shell, regardless of atomic number. They compensate for an increased positive charge in the nucleus by filling the partially occupied *f* subshell of the fourth shell. It is because the valence-electron shell of the rare-earth atoms remains unaltered that they have much the same chemical properties.

What differences there are in the chemical behavior of the rare earths are largely the consequence of the different effective atomic sizes the elements exhibit when they shed three electrons to become trivalent ions. Owing to the progressively stronger attractive force between the increasing positive charge in the nucleus and the increasing negative charge in the 4*f* subshell, the size of the trivalent ions shrinks with increasing atomic number. Important exceptions to the pattern in ionic size occur when certain rare-earth ions adopt "anomalous" ionic states, different from the usual trivalent state. Under reducing conditions (in the presence of elements that easily shed electrons) europium may gain an electron and assume a divalent state; under oxidizing conditions (in the presence of free oxygen) cerium may lose an electron to become quadrivalent.

In a rock melt or hot aqueous solution these ionic-size differences determine which rare-earth elements will be preferentially incorporated into crystals that form as the liquid cools. The reason is that a rare-earth atom is usually not a free ion but is caged by other ions, called ligands, in an arrangement known as a complex. The number of ligands in a complex as well as their geometric arrangement depends on the concentration and species of ligands available in the liquid. Before it can be incorporated into a crystal, a complex bearing a rare-earth ion must be somewhat distorted at the crystal surface—a process that is critically dependent on the electrochemical conditions at the solid-liquid interface. Once the complex (including the rare-earth ion) is incorporated into the thin surface layer, it has to be rearranged to conform to the latticework of the growing crystal. In the event that the complex is not assimilable, the rare-earth ion can nonetheless find its way into a growing crystal, if the complex can disassemble itself so that the ion is freed at the crystal surface.

Xenotime and monazite, for example, are both phosphate minerals that can crystallize from chemically similar melts having rare earths as a minor constituent. Yet the crystal structure of xenotime provides only enough room for rare-earth complexes that have a maximum of eight oxygen ligands, whereas the crystal structure of monazite allows a rare-earth atom in a complex to be surrounded by as many as 10 oxygen lig-



VARIOUS GEOLOGIC PROCESSES can yield minerals enriched in rare earths. Magmas (melted rock) can have high concentrations of rare-earth elements—particularly the "lighter" ones (those having lower atomic numbers)—if they are produced in carbon dioxide–rich regions deep in the upper mantle. Such a magma can occasionally rise to the crust, where it forms bodies known as carbonatites. Mantle magmas produced closer to the crust, where carbon dioxide and other volatile substances have escaped to the earth's surface, are generally depleted in rare earths. Crustal material can also melt to produce granitic magmas that have a higher rare-earth content than the solid rock from which they originate. As such magmas cool, crystals form that preferentially incorporate "heavy" rare-earth elements (those having higher atomic numbers). The liquid residue after extensive crystallization can therefore be highly enriched in light rare-earth elements. The residue finally crystallizes into rocks known as pegmatites. Rare earths in a hydrothermal, or hot-water, solution can also be concentrated as the solution crystallizes. In this case, however, it is the heavy rare earths that become progressively more concentrated in the liquid residue.

ands. Because the smaller rare-earth ions are surrounded by fewer oxygen atoms, they can occupy the smaller vacancies in a growing crystal of xenotime. Larger rare-earth ions, in contrast, are surrounded by more oxygen atoms, and they can occupy the larger open spaces of the monazite crystalline lattice. As a result monazite tends to incorporate "lighter" rare-earth elements, which have lower atomic numbers, whereas xenotime tends to incorporate "heavier" rare-earth elements, which have higher atomic numbers. Most minerals have crystals that accommodate complexes with only six oxygen ligands and thus show a strong affinity for the heavier rare earths.

Although the size of the rare-earth ions also determines which ion is preferentially incorporated into crystals in hot aqueous solutions, the end result is the reverse: the heavier, smaller rare-earth elements tend to remain in the liquid phase, whereas the lighter, larger rare-earth ions tend to be contained in crystals. The reason is that in an aqueous solution rare-earth ions are surrounded mainly by polarized water molecules rather than ligands. The tight sheath of water molecules allows only a few negatively charged ions, such as hydroxyl (OH^-), fluoride (F^-) or carbonate (CO_3^-) ions, to come in contact with the rare-earth ion in order to establish ligand bonds. The strength of such bonds depends on the charge density of the rare-earth ion. Because in most circumstances every rare-earth ion acquires the same charge of +3, the charge density of a rare-earth ion is in effect a function of its size: smaller ions have a higher charge density than larger ones. It therefore requires less energy for a larger, lighter rare-earth ion to shed complexing ligands so that the ion can become assimilated into a crystalline lattice.

The first application of such principles of physical chemistry to geology was undertaken by a Swiss-born Norwegian mineralogist and petrologist, Victor M. Goldschmidt. To explain the incorporation of elements into minerals formed from magma (melted rock) and hydrothermal (hot water) solutions, he formulated in 1937 a set of rules based on the similarity of ionic size and charge between crystal-forming elements. Goldschmidt's concepts are too simplistic, however, because they disregard the role of complex formation. Only now are geologists taking into account the advances made in the study of complexation to help explain how the rare-earth elements were able to fractionate, or segregate themselves according to atomic number, throughout the earth's crust and mantle, given the elements' initial abundance at the time of the earth's formation about 4.5 billion years ago.

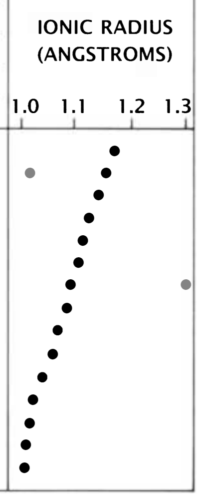
The abundance of the rare earths in the primordial earth can be deduced from the elemental composition of a special group of meteorites known as carbonaceous chondrites. Both the earth and carbonaceous chondrites are thought to have been formed from material that condensed and accreted from a primeval cloud of gas and dust. Evidence for the common provenance of the material was contained in measurements taken in 1976 by Gerald J. Wasserburg and Donald J. DePaolo of the California Institute of Technology.

Wasserburg and DePaolo measured the amount of radioactive samarium 147 (^{147}Sm) and its daughter product neodymium 143 (^{143}Nd) in meteorites and terrestrial rocks. Because of the radioactive decay of samarium 147, the amount of neodymium 143 has increased steadily over eons in relation to neodymium's stable isotopes, such as neodymium 144, at a rate determined by the original ratio of samarium to neodymium in the material and by the half-life of samarium 147 (approximately 100 billion years). Wasserburg and DePaolo were able to demonstrate that both the earth (taken as a whole) and the chondritic meteorites have the same $^{143}Nd/^{144}Nd$ ratio and therefore must have had similar initial Sm/Nd ratios. Absolute concentrations of samarium and neodymium, however, in the primordial mantle are thought to have been approximately twice those in the meteorites.

The reason for this is grounded in the observation that rare-earth elements reside almost exclusively in the silicate fraction of the meteorites rather than the metallic fraction. Because of the elements' affinity for silicates, it seems reasonable to conclude that they were partitioned (and thereby slightly concentrated) into the primordial mantle as the metals migrated to the core of the earth.

Samarium and neodymium, as well as the other rare earths, were further concentrated in the outer, crustal layers of the earth once the young mantle built up enough regional heating to undergo partial melting. The reason is that the first liquid exuded from a heated rare-earth-bearing mineral tends to incorporate the lighter rare earths, which form the stabler complexes in the liquid state. Hence as the material of the young

ATOMIC NUMBER	ELEMENTS		ELECTRON CONFIGURATION OF THE OUTER SHELLS: 4 (s p d f: 2 6 10)	5 (s p d: 2 6)	6 (s)
57	LANTHANUM	La	0	1	2
58	CERIUM	Ce	1	1	2
59	PRASEODYMIUM	Pr	3	0	2
60	NEODYMIUM	Nd	4	0	2
61	PROMETHIUM	Pm	5	0	2
62	SAMARIUM	Sm	6	0	2
63	EUROPIUM	Eu	7	0	2
64	GADOLINIUM	Gd	7	1	2
65	TERBIUM	Tb	9	0	2
66	DYSPROSIUM	Dy	10	0	2
67	HOLMIUM	Ho	11	0	2
68	ERBIUM	Er	12	0	2
69	THULIUM	Tm	13	0	2
70	YTTERBIUM	Yb	14	0	2
71	LUTETIUM	Lu	14	1	2



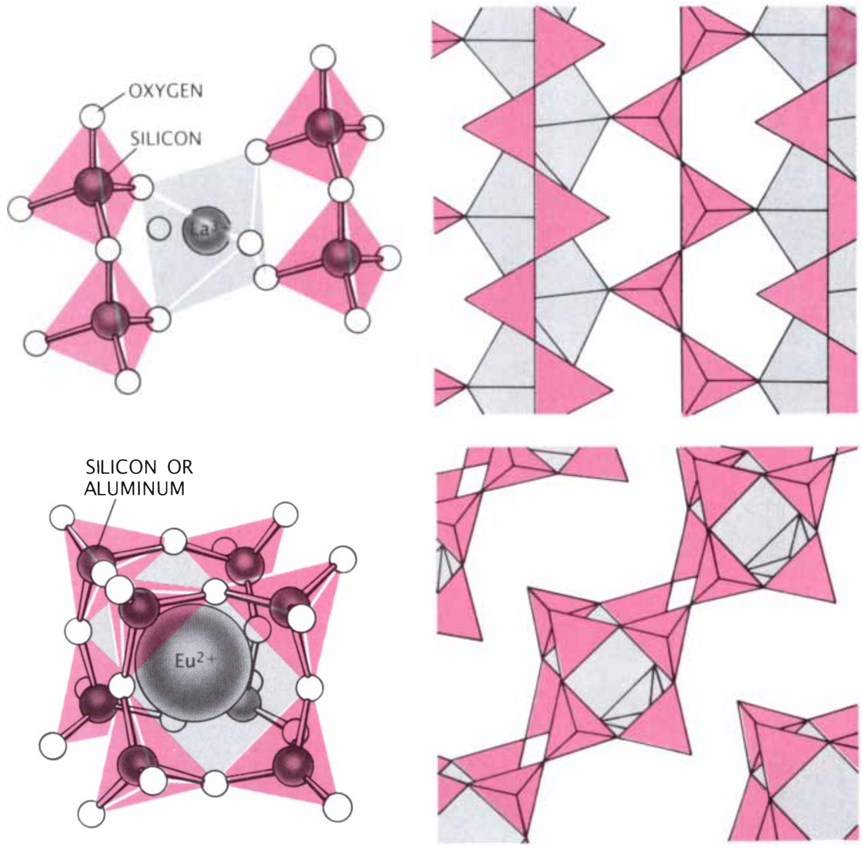
RARE-EARTH ELEMENTS include the lanthanides: the elements that have atomic numbers 57 through 71. Atoms of all the rare earths display very similar chemical behavior, because the valence, or bonding, electrons in their outer "shells" have similar quantum states (designated by the letters *s, p, d* or *f*). Although the configuration of the valence electrons does not change significantly with increasing atomic number, the ions (produced when the atoms shed three electrons) exhibit a gradual and systematic decrease in size. (Because the ions constituting a crystal can be visualized as charged spheres, they can be assigned a "radius" corresponding to their size.) Cerium and europium ions deviate considerably from the smooth ionic-radius trend, when they respectively lose or gain an electron (*red*)—as they sometimes do in underground environments.

mantle began to melt, the rare-earth elements were partitioned, the larger and lighter elements (such as samarium and neodymium) being incorporated in the melt more than those that are smaller and heavier. Since solids are generally denser than liquids, these early melts had a lower density than the solid parent material in the mantle and rose toward the surface of the earth to form the first primitive crust.

This type of elemental segregation was already well advanced 3.8 billion years ago, during the Archean era, when sediments were deposited to form the oldest terrestrial rocks known. Such rocks can provide a reliable estimate of the rare-earth concentrations at that time, since rare-earth solubilities in surface waters are extremely low and the rare earths are not readily extracted from their host mineral by weathering, erosion, transportation or redeposition. The rocks reveal that the Archean upper crust indeed had an enrichment factor in relation to the early mantle of about 25 for lanthanum, decreasing smoothly with atomic number to a factor of about five for lutetium.

The separation of the continental crust from the early mantle is thought to have reached a high point between two and 2.5 billion years ago. As much as 75 percent of the current volume of the continental crust may have been formed by that time. The extraction of such large volumes of rare-earth-enriched material left some regions of the upper mantle depleted in the lighter rare-earth elements. Because neodymium is lighter than samarium, the Sm/Nd ratio (which determines the rate of increase of the measurable $^{143}Nd/^{144}Nd$ ratio) was raised sufficiently in the depleted mantle to endow it with a distinctive isotopic signature. Indeed, the high $^{143}Nd/^{144}Nd$ ratios of volcanic material collected along mid-oceanic ridges reveal that the material's mantle source probably was involved in the early phases of continent formation.

In spite of the early mantle's general depletion in light rare-earth elements, some mantle magmas that are extruded through the crust of the earth nonetheless have elevated concentrations of rare-earth elements. How is this possible? The answer lies in the fact that minerals having high melting points can crystallize out of a mantle magma and be physically removed from it by gravitational settling or flotation. Since the rare-earth



RARE-EARTH IONS in a silicate melt are surrounded by other ions, called ligands, forming what is known as a complex. The structure of a rare-earth complex depends on the concentration and type of ligands available; it also depends on the size of the rare-earth ion. A rare-earth element is most likely to be assimilated into a crystal growing in the liquid if its complex happens to have a structure that is congruent with the crystal's basic lattice structure; otherwise the complex has to be disassembled. For this reason a lanthanum ion (which has a charge of +3) tends to be incorporated into silicate minerals containing iron and magnesium (*top*), whereas a europium ion (which has a charge of +2 and is much larger) is preferentially incorporated into feldspars (*bottom*).

elements (in particular the lighter ones) usually tend to stay in the liquid phase, an enrichment in rare-earth elements results as the magma becomes crystallized. Indeed, extensive crystallization can yield a hundredfold increase in rare-earth concentration in mantle magmas. Nevertheless, the rare-earth content of the rocks formed from such magmas (which can be found in oceanic islands and continental-rift zones) is still considerably below the content required to yield minable ores.

Fortunately, under favorable conditions mantle magmas are subjected to another source of rare-earth enrichment that does yield minable ores. These conditions correspond to the formation of carbonatites (igneous carbonate rocks). Carbonatitic magmas are produced when mantle rocks melt in the presence of large amounts of carbon dioxide. In such magmas the rare-earth elements readily form strong carbonate complexes, although the lighter elements form stronger complexes than the heavier elements.

In carbonatitic magmas rare-earth levels can be high enough to crystallize such rare-earth minerals as bastnaesite and parisite. In such a case the rocks formed as the magma cools can show individual rare-earth levels some 10,000 times higher than the early mantle, and they contain minerals having a total rare-earth-oxide concentration of as much as 60 percent. Carbonatites constitute the major source of rare-earth ores, an example of which is the Mountain Pass deposit in California.

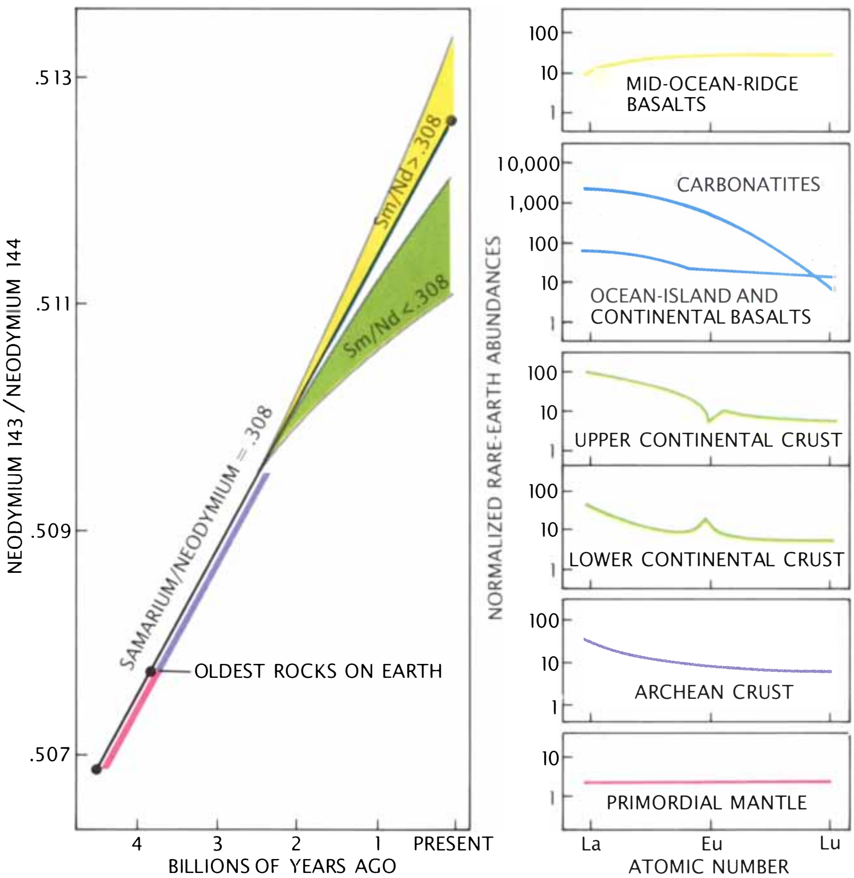
As the continental crust became thicker, the temperature in its lower layers became high enough to cause localized melting. Eventually the melting and the subsequent redistribution of the liquid material within the crust became the dominant fractionating geologic process—more so than the accumulation of mantle-derived magmas.

Magmas produced in the crust are considerably richer in silica (silicon oxide) and alumina (aluminum oxide) than mantle melts are. In such a melt rare-earth ions are enclosed in complexes composed of silicon, aluminum and oxygen. Aluminosilicate minerals that crystallize from such melts will incorporate groups of complexes that conform to their crystal structure. Yet owing to the high electric charge of a rare-earth ion, an aluminosilicate complex that encloses a rare-earth ion is deformed and often cannot readily serve as a crystal "building block." Hence complexes that contain rare earths tend to accumulate in the liquid phase of a crystallizing crustal melt.

Yet a rare-earth ion can also be incorporated in aluminosilicate crystals, if the complex is first disassembled. Since the more deformed complexes have weaker ligand bonds, it takes less energy to break them up. The extent of the complex's deformation is determined largely by the ion's charge density; the heavier rare earths (which have smaller ionic sizes) deform the complexes more than the lighter ones. Consequently the heavier rare earths are more easily incorporated into an evolving crystal structure, leaving the residual melt enriched in the lighter rare earths.

If a crustal melt has enhanced rare-earth concentrations in the first place and undergoes considerable crystallization, the residual liquid becomes strongly enriched in the light rare earths as well as in water and fluorine, which are also not readily incorporated into crystals. When the residual liquid finally crystallizes, it does so as relatively small bodies of very coarse-grained rocks, known as pegmatites, that can contain substantial accumulations of rare-earth minerals and provide a further source of industrial quantities of the rare-earth elements—the light ones in particular. Examples of such deposits can be found in the South Platte district in Colorado.

Stuart Ross Taylor and Scott M. McLennan of the Australian National University have shown that sediments older than about 2.5 billion years, which represent a sampling of the Archean continental crust, are indeed poorer in rare-earth elements than their younger equivalents. The younger sediments, however, show a small but distinct depletion in europium. The discrepancy stems from the fact that plagioclase, a calcium-rich feldspar that is often a residue of crustal melts, preferentially retains europium in its anomalous, divalent form. As a result intracrustal melting has led to enrichment of the upper crust in rare-earth elements—specifically the light ones—but a relative depletion in europium.



FRACTIONATION, or segregation of elements according to atomic number, is a by-product of the geologic evolution of the earth, as is indicated by the ratio of neodymium 143 (^{143}Nd) to neodymium 144 (^{144}Nd) in various types of rock (*left*). Because neodymium 143 is produced by the radioactive decay of samarium 147, the ^{143}Nd/^{144}Nd ratio reflects the ratio of samarium to neodymium. The primordial earth is assumed to have a Sm/Nd ratio of .308, since this is the ratio in chondritic meteorites. The abundance of each rare earth in such meteorites is taken as the base value from which enrichment factors, known as normalized abundances, are calculated. The curve connecting normalized rare-earth abundances (*right*) is therefore assumed to be flat for the primordial mantle, since it has not had an opportunity to fractionate. Rocks formed on the Archaen crust, which are the oldest rocks known on the earth, already show enrichment in rare-earth elements—particularly the lighter ones. Rocks that are less than 2.5 billion years old have ^{143}Nd/^{144}Nd ratios that betray conspicuously different Sm/Nd ratios, reflecting the fact that the melting and crystallization leading to their formation either enriched or depleted them in the lighter rare-earth elements. The prominent discrepancies in the abundances of europium in rocks from the upper and the lower continental crust result from the fact that the lower crust preferentially retains europium.

Once they have risen to the upper continental crust, liquid or solidified magmas may interact with underground hot water to produce hydrothermal solutions. Such a hydrothermal solution can crystallize into minerals whose rare-earth content betrays the elements' original source and the conditions prevailing during the minerals' deposition. (Interpretations of the data, however, often must be tentative because circulating hydrothermal solutions have ample opportunity to modify their chemistry as they interact with the surrounding rocks.)

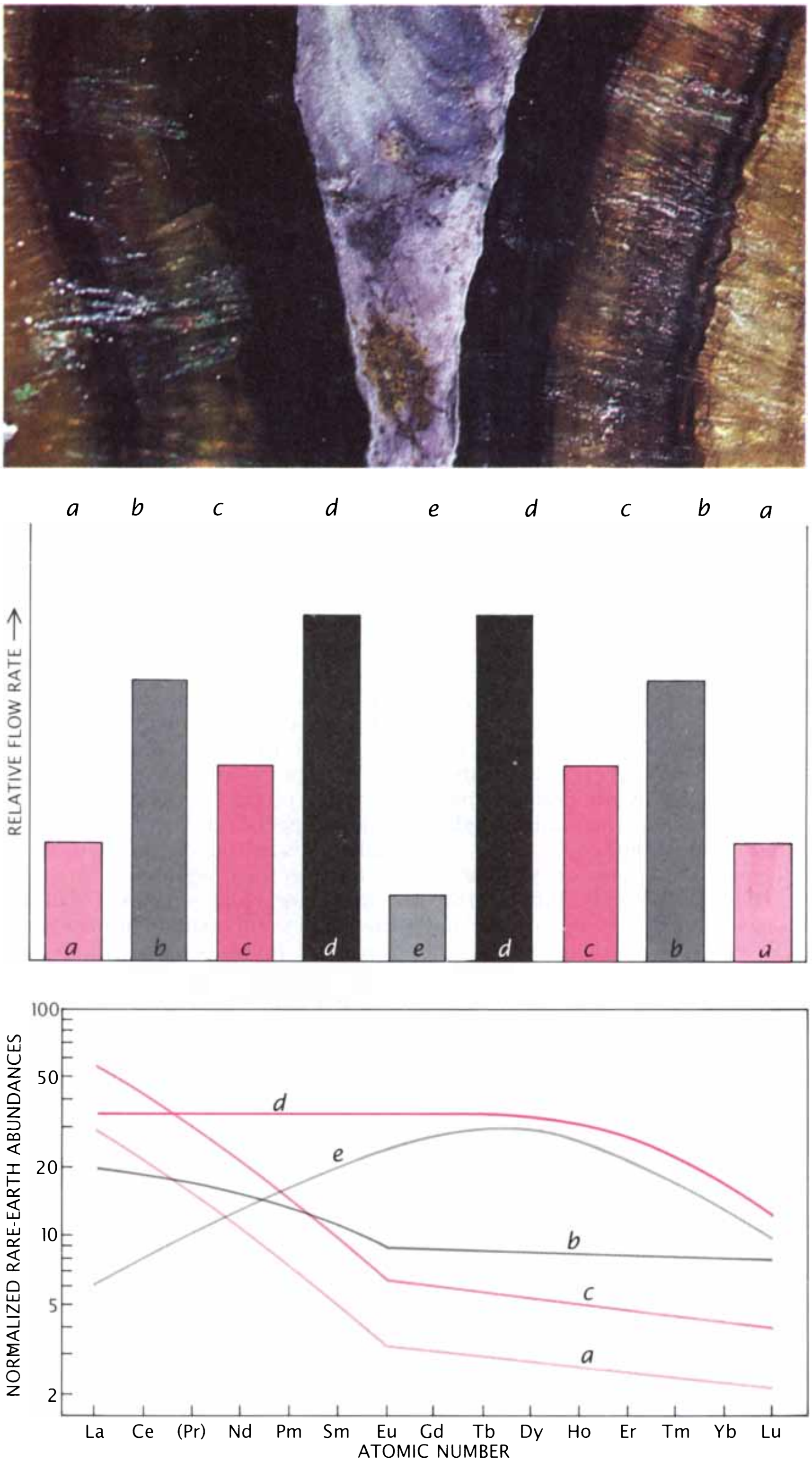
The rare-earth elements are most effectively deposited in calcium minerals, such as fluorite and calcite, since the trivalent rare-earth elements tend to substitute for calcium.

Because the heavier rare-earth elements form the stabler complexes in aqueous solutions, light rare earths have a slightly greater chance of being incorporated into the crystals that grow in hydrothermal solutions. As such a solution crystallizes, it yields minerals with progressively higher proportions of heavy rare-earth elements. Yet calcium minerals are so effective at incorporating all sizes of the trivalent rare-earth ions that the enrichment of heavy rare earths to a concentration in which they could crystallize to form their own minerals rarely occurs. As a consequence rare-earth elements from hydrothermal ore deposits are recovered as by-products only.

When the role of chemical complexation of the rare-earth elements in hydrothermal solutions is compared with that in magmatic melts, it is clear they follow opposing trends. In magmatic systems the light rare earths show a strong tendency to be retained in the liquid phase, whereas in hydrothermal solutions it is the heavy rare earths that show such a tendency. Considered in the context of geologic time, the opposing tendencies result in an overall diminished fractionation of the rare earths in the crust of the earth.

The opposing fractionation processes explain the relative paucity of rare-earth-ore deposits, in spite of the fact that the rare earths are not at all that scarce (in global terms). Lead is less abundant on the earth than the rare earths (taken as a group), yet its ores can frequently be found in rich lodes, because in lead's case magmatic and hydrothermal fractionation processes reinforce one another.

Nevertheless, the study of the distribution of rare earths in minerals has shown that under favorable circumstances (those in which either magmatic or hydrothermal processes clearly dominate) final concentrations can be high enough to form viable deposits of rare-earth ores. Moreover, such investigations also shed light on the way elements are exchanged between minerals and a liquid consisting of either a magma or a hydrothermal solution. Because geochemists generally agree that such processes are the principal means by which most elements have been concentrated in the rocks of the earth's crust and mantle since primordial times, the study of the distribution of rare earths in minerals helps to ensure the continuity of all mineral supplies—not just rare-earth ores—for future generations.



HISTORY OF HYDROTHERMAL MINERALIZATION is revealed by the pattern of rare-earth abundances in symmetrical bands of fluorite that lined the walls of a rock fissure (*top*). In contrast to magmatic mineral formation, it is the heavy instead of the light rare-earth elements that are preferentially retained in the liquid phase. Hence as a hydrothermal solution crystallizes, the liquid residue is progressively enriched in the heavy rare-earth elements. The fluorite bands reflect this behavior, which is mediated by the flow rate of the solution (*middle*). The bands on the outside, which were the first ones formed, have a markedly higher concentration of light rare earths than of heavy rare earths, as can be seen from the plots of normalized abundances (*bottom*). The heavy rare earths in turn are preferentially incorporated into the fluorite bands on the inside.

Art, Illusion and the Visual System

Form, color and spatial information are processed along three independent pathways in the brain. That explains why certain images can create surprising visual effects

by Margaret S. Livingstone

Seeing is much more complicated than most people realize. In this age of video films and television it is tempting to think of vision as just another way of making a picture, but cameras can only record what they see; they cannot interpret or identify the images they create. No video system or computerized camera, no matter how sophisticated, can match the ability of the human visual system to make sense of an infinite variety of images. That ability is made possible by the brain's capacity to process huge amounts of information simultaneously.

Recent findings, including some by my colleague David H. Hubel and me, suggest that visual signals are not processed by a single hierarchical system but are fed into at least three separate processing systems in the brain, each with its own distinct function. One system appears to process information about shape perception; a second, information about color; a third, information about movement, location and spatial organization. It is now likely that some of the visual effects created by artists and designers succeed because light information is analyzed in this way.

The notion that the visual system processes information along several separate pathways is far from new; it probably dates back to the mid-1800's, when scientists first observed that the optic nerve splits into several subdivisions after leaving the eye. It may seem strange to think of vision as a multipartite rather than a single process—strange to think that perception of the shape of an object, its color and where it is and how it is moving are processed by different parts of the brain—yet anatomical, physiological and psychological studies strongly indicate this is the case. The fact that an object's shape, color, position and motion appear unified even though each component is analyzed separately can be compared to the experience of listening to someone speak. You hear that person's voice and see his mouth move without being aware that the two are processed independently.

The visual process begins when light passes through the lens of the eye to the retina, a thin sheet of highly specialized neural tissue at the back of the eye. There the light strikes special photoreceptor cells, called rods and cones, and is converted into electrical signals.

Rods are more sensitive to light than cones and respond only to very low light levels. Cones, which operate at higher light levels, can be divided into three types, which contain different pigments (light-absorbing molecules). The cone pigments absorb light over a broad range of the visible spectrum, but each type of pigment is maximally sensitive to a different part of the spectrum. One type of cone is most responsive to short wavelengths (the blue and green region of the spectrum), the second type to middle wavelengths (green light) and the third to still longer wavelengths (yellow, orange and red light).

In subsequent stages of color analysis, the color selectivity of the signals is increased by comparing, that is, subtracting, the signals from the three cone types. Since the strength of a cone's signal is proportional to the light intensity over a rather broad range of wavelengths, the strength of that one signal does not reveal much about the color of the light, but comparing the signals received from two different types of cones does. When strong signals from red cones are measured against weaker signals from green and blue cones, for example, one sees red; equal signals from the red and green cones and weak signals from the blue cones, on the other hand, translate into yellow light, and so on. The comparison of wavelengths is made by neurons whose signaling rate is speeded up by signals from one kind of cone and slowed down by signals from another kind. The net effect is essentially a subtraction of one cone type from another.

When light levels are too low for detection by the cones, as is the case at night, the more sensitive rods are stimulated. Because there is only one kind of rod, however, different wavelengths cannot be compared and people therefore are completely color-blind in dim light.

Before passing out of the eye and into the brain, the electrical signals from the photoreceptor cells are processed by a second stage of neurons and then transmitted to an inner layer of retinal cells called ganglion cells [see "The Functional Architecture of the Retina," by Richard H. Masland; SCIENTIFIC AMERICAN, December, 1986]. The ganglion-cell layer represents the first major subdivision in the visual pathway. This layer contains two intermixed cell types that differ in size (they are large and small) and in the way they process the information received from the cone cells. The large cells do not distinguish one type of cone-cell signal from another; they simply add the information received from the three types. Because they lack color-selectivity, they can be thought of as color-blind.

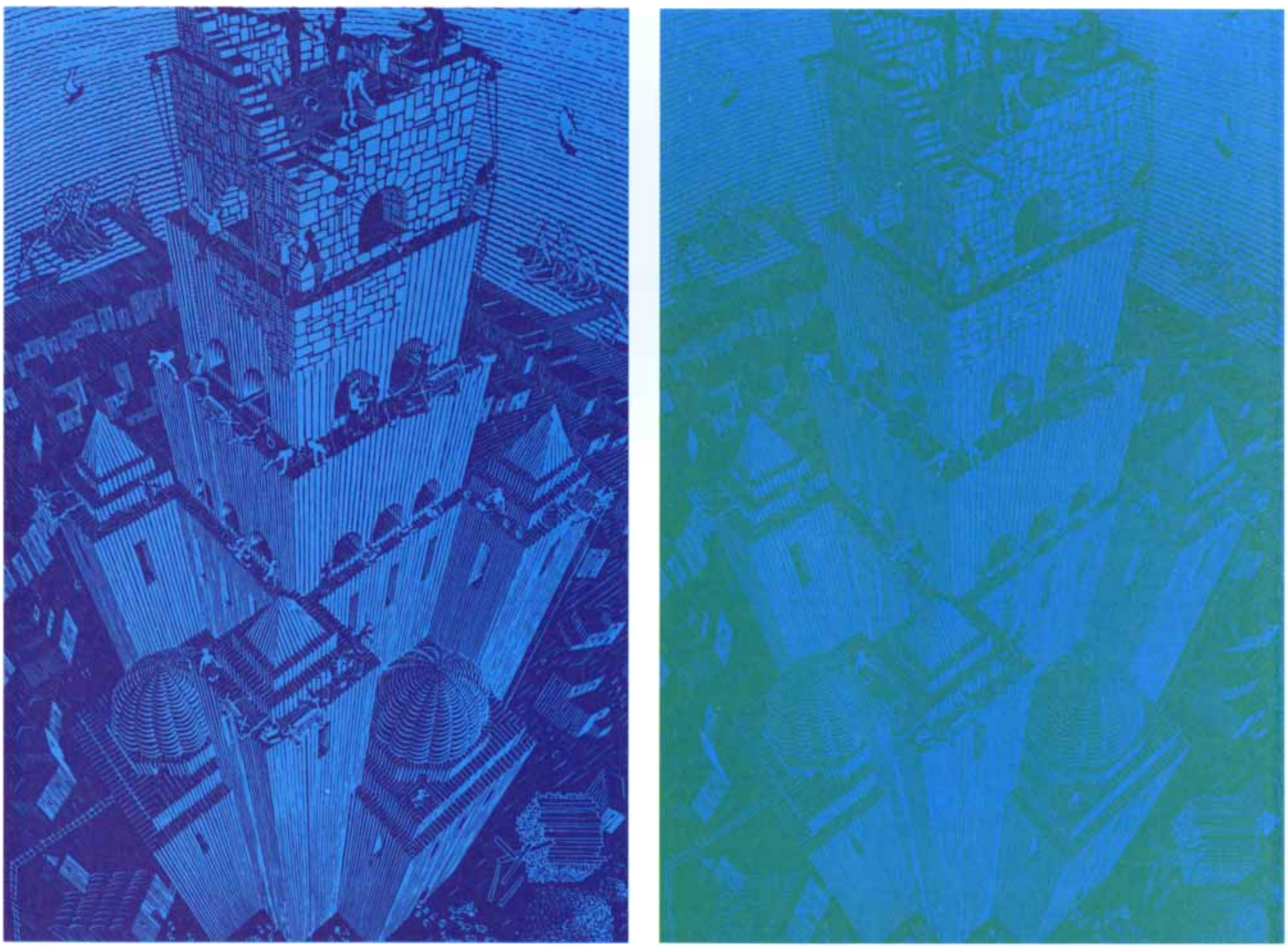
The small ganglion cells, on the other hand, do distinguish between the three cone-cell types and in effect subtract information received from them. This enables them to signal information about different colors. A red-minus-green ganglion cell, for example, will respond only to red light, although it may receive information from several types of cone cells. The result is that the signals from these small ganglion cells are more color-selective than the input signals they receive from the cone cells.

The ganglion cells transmit their signals via the optic nerve from the eye to the lateral geniculate bodies, two peanut-size clusters of neurons deep in the brain. The lateral geniculate bodies, like the ganglion-cell layer, consist of two classes of neurons that differ from each other both in size and in the kind of information they signal. Unlike the ganglion layer of the retina, however, these two classes of cells are not intermixed but are segregated into spatially distinct subdivisions. The small-cell (parvocellular) division of the geniculate receives input from the small ganglion cells and the large-cell (magnocellular) division receives input from the large ganglion cells.

Work in a number of laboratories indicates that these two pathways, called the parvo and magno systems, differ not only in their color selectivity but also in their contrast sensitivity, temporal resolution and acuity. The magno system is more sensitive to brightness contrast and has a faster response time and lower acuity than the parvo system. These striking differences in the visual-response properties of the two subdivisions suggest that they serve quite different functions, but to understand them it is necessary to consider them in the context of the higher levels of the visual system.

From the lateral geniculate bodies, nerve signals pass to the first cortical visual area, visual area 1. This area is a folded sheet of neurons the size of a credit card (but three times as thick) situated at the very back of the brain. The path of electrical impulses can be traced to the middle layer of visual area 1, where the segregation of visual information is maintained. Signals from the magno system are channeled into the top half of this layer and signals from the parvo system into the bottom half.

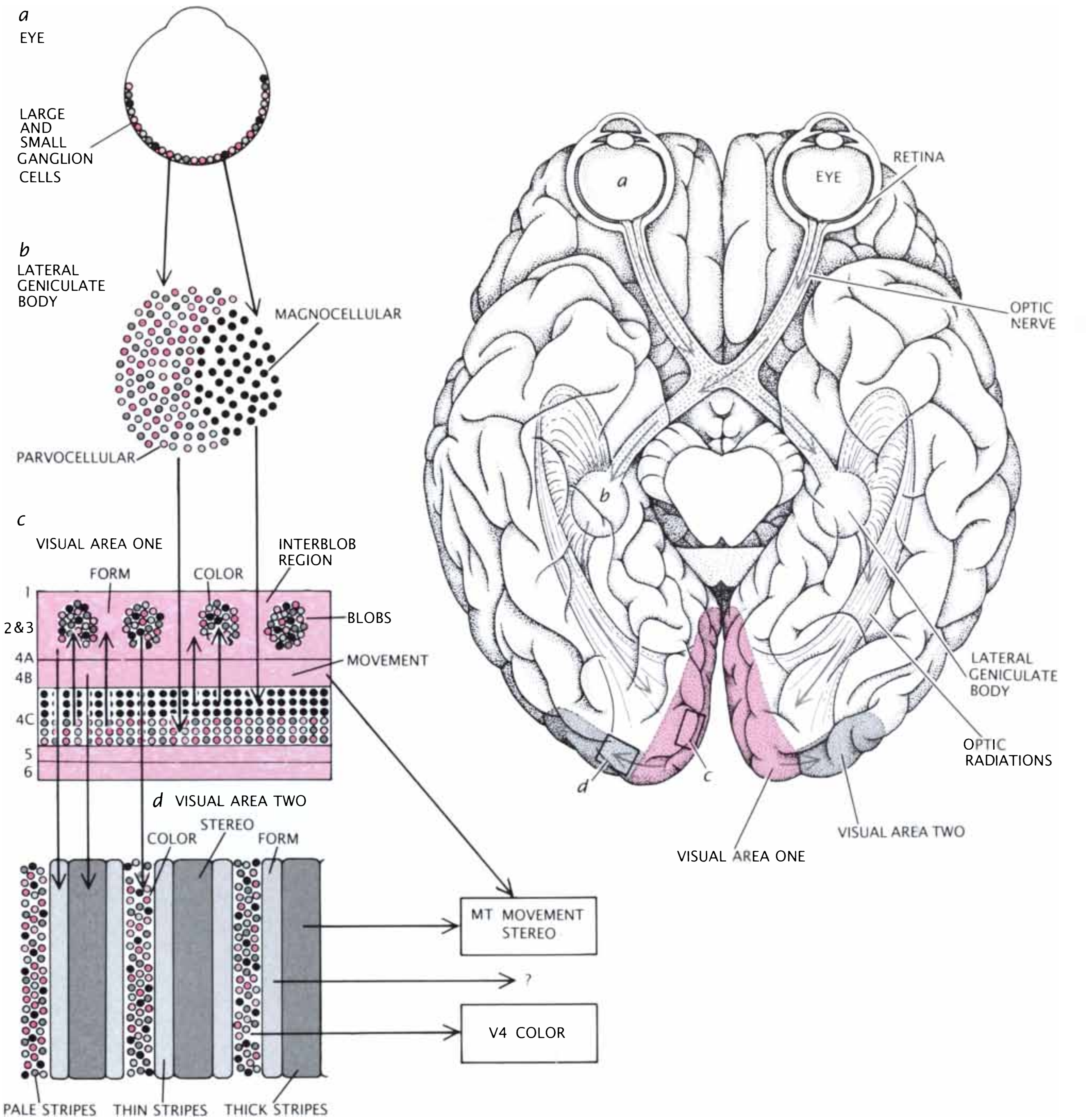
The stage that follows this prominent functional segregation of information has been clarified only recently. Investigators realized that information must be processed along separate pathways at still higher stages in the brain, because segregation of different visual functions had been discovered in regions beyond visual area 1. Semir Zeki of University College London found an area in the middle temporal lobe in monkeys, for example, that has a high proportion of cells sensitive to movement or stereoscopic depth. He also



TOWER OF BABEL, by M. C. Escher, demonstrates the effect of luminance contrast on depth perception. Colors with different luminance levels, in this case blue and black (*left*), create the perception of a vividly three-dimensional image. But when the black is replaced (*right*) with a shade of green that has luminosity close to that of blue, the three-dimensional effect is lost and the image is hard to see. To convince yourself that both images are identical, look at them through a piece of blue glass or plastic.

found a second area, called visual area 4, that seems to be selectively involved in color perception. The relation between functional segregation in these higher visual areas and the subdivisions of the geniculate bodies until recently was not known. Research that Hubel and I have carried out at these intermediate levels has enabled us in a sense to complete the puzzle.

A critical piece of information was provided in 1978 when Margaret Wong-Riley of the University of California at San Francisco found that staining visual area 1 for a mitochondrial enzyme, cytochrome oxidase, revealed a fine pattern of dark and light regions in its upper layers. The dark regions turned out to be slightly irregular ovoids, about .2 millimeter



HUMAN VISUAL SYSTEM consists of three distinct pathways. Light entering the eye strikes the retina (*A*), where it is converted into electrical impulses that travel to the brain along the optic nerve. The first split in information processing is apparent in the lateral geniculate bodies (*B*), where the small cells of the parvo system carry information about color contrast and the large cells of the magno system carry information about luminance contrast. From the magno cells information is sent to layer 4B of visual area 1 (*C*) and then to the thick stripes in visual area 2 (*D*). There the signals are analyzed to give information about motion and depth. Input from the parvo system is sent to the interblobs of visual area 1 and then to the pale stripes in visual area 2, where it is analyzed for information about shape. Input from both the parvo and the magno systems is combined in the blobs, where it is processed for color and luminance. It then passes to the thin stripes of visual area 2 and from there to visual area 4.

in diameter and arranged in a semiregular mosaic. We decided to call the ovoids blobs (because of their shape) and the lighter area surrounding them interblobs. Once we had identified these subdivisions it was possible to look at the connections and response properties of each subdivision. The two pathways evident at lower levels seem to rearrange themselves into three subdivisions at this stage: the interblobs receive input from the parvo system, layer 4B receives input from the magno system and the blobs seem to receive input from both.

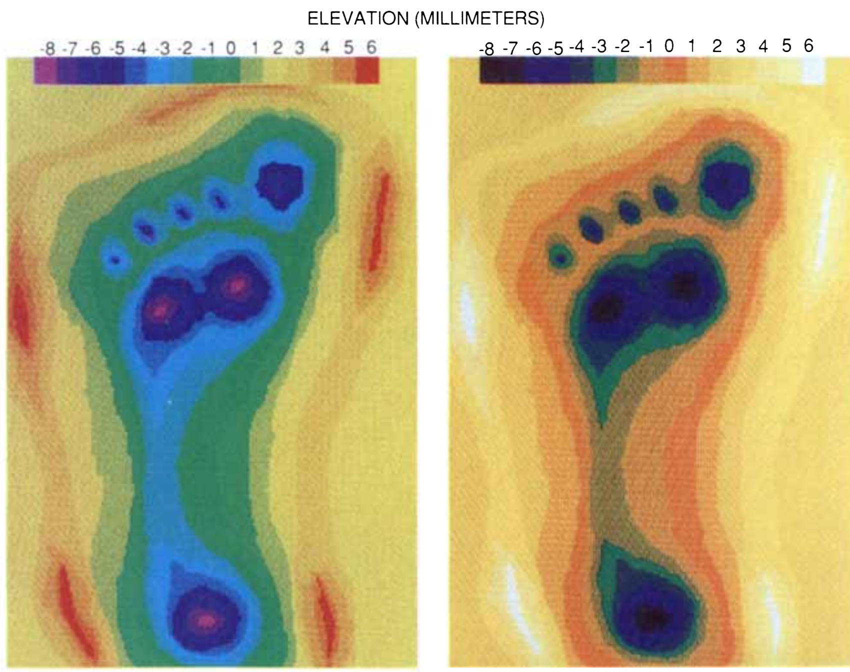
We began by testing the visual response of cells in each of the three subdivisions of visual area 1, asking which stimulus variables were most important in each of the three subdivisions. We measured the selective responses of the different neurons to shape, position, distance, movement, color, brightness and size and found that the differences among these neurons were dramatic. The blobs contain cells that are highly selective for color or brightness but are not at all selective for shape or movement. Interblob cells are selective for orientation but not for color or movement. An interblob cell may respond to a vertical bar, for instance, regardless of how it moves or whether it is black, white or colored—the only criterion is that the bar be vertical; that same cell will not respond to bars in any other orientation. Cells in layer 4B are also unselective for color, but they are selective for orientation and movement; a cell in this system, for example, will respond either to horizontal bars that move upward or to vertical bars that move horizontally but not to both.

Next we turned our attention to visual area 2, a region of the cortex adjacent to visual area 1 that receives input from the three pathways of visual area 1. We applied the same mitochondrial stain to visual area 2 and found this area is also differentiated into three subdivisions. Here, however, the subdivisions are characterized by three kinds of alternating stripes that we call pale stripes, thin stripes and thick stripes because of the way they stain.

We measured the response of cells in the different stripes to various visual stimuli and obtained the following results. The color-selective blobs of visual area 1 provide input to the thin stripes of visual area 2, which continue the processing of color information. The orientation-selective interblobs provide input to the pale stripes, which process the information in a way that suggests they are involved in shape analysis. The magno system provides input to the thick stripes, which analyze information about stereoscopic depth.

Zeki and his colleague Stuart Shipp, and John Maunsell and David C. Van Essen of the California Institute of Technology, have accumulated evidence that indicates the three pathways continue to remain separate at still higher stages of the

SHADOWS can be any color; they need only to be darker than the rest of the surface to convey a sense of depth. This is demonstrated here in this reproduction of a self-portrait by Henri Matisse. The green color used to shadow the face (*left*) is peculiar, but the three-dimensional effect it creates seems normal. In the black-and-white photograph of the painting (*right*) it can be seen that the shadows are indeed darker than the face.



CONTOUR MAPS are often in color, which makes it easier to correlate any given point on the map with a scale value. Most contour maps use color scales based on the visible spectrum, with red representing high and violet low values (*left*). The disadvantage of such a scale is that the shape of the object may not be readily apparent. That problem can be overcome (*right*) with a color scale based on brightness, where the degree of brightness corresponds to elevation. When a contour map is based on brightness, it is possible to discriminate values as well as the overall shape of the object at a glance.

brain's cortex. They have found that the thick stripes project to the middle temporal area (MT), a region concerned with movement and stereopsis (the ability to judge depth based on the differences between the images in the two eyes), and the thin stripes to visual area 4 (V4), an area concerned with color, but they are not yet certain of the pathway of projections from the pale stripes.

Further evidence that color, movement and the perception of form are carried by separate pathways comes from clinical neurology. Localized brain damage from strokes can produce highly selective types of blindness, such as loss of the ability to recognize faces, or loss of color vision without loss of the perception of form. Cortical color blindness is a rare syndrome resulting from damage to an area of the brain that may correspond to visual area 4. Because the blob system carries information not only about color but also about brightness, patients with damage to their color system can be expected to show defects in their ability to differentiate between shades of gray as well as in their perception of colors. Indeed, one such patient complained that colors were washed out and that freshly fallen snow looked gray and dirty to her.

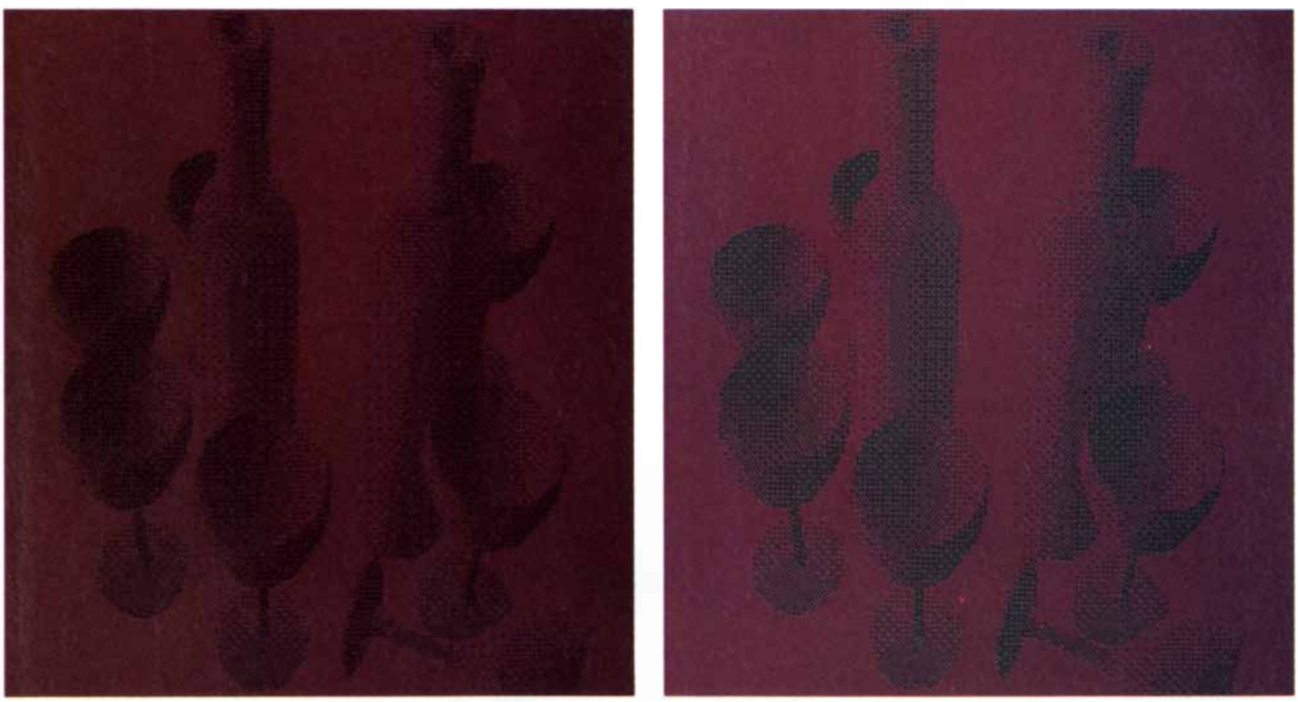
Based on these physiological and anatomical findings it is possible to summarize the probable functions of the three subdivisions of the visual system as follows:

The parvo-interblob-pale-stripe system carries high-resolution information about borders that are formed by contrasting colors. Although neurons in the early stages of this system are color-selective, those at higher levels respond to color-contrast borders but do not carry information about what colors form the border. Because much of the information about the shape of objects can be represented by their borders, we suspect that this system is also important in shape perception. The slow time course and high resolution of this system are probably important for the ability to see stationary objects in great detail.

The blob-thin-stripe-V4 system processes information about color and shades of gray but not about movement, shape discrimination or depth. This system has a severalfold lower acuity than the interblob system and therefore sees objects in color but not in great detail.

The magno-4B-thick-stripe-MT system carries information about movement and stereoscopic depth. Neurons in this system have a very fast response time, but their responses decay rapidly even when the stimulus is maintained, so that the system is particularly sensitive to moving stimuli. Thus it is good at detecting motion but poor at scrutinizing stationary images. In addition it appears to be color-blind: it is unable to see borders that are visible only on the basis of color contrast.

Cells in the parvo system can distinguish between red and green at any relative brightness of the two. Cells in the color-blind magno system, on the other hand, are analogous to a black-and-white photograph in the way they function: they signal information about the brightness of surfaces but not about their colors. For any pair of colors there is a particular brightness ratio at which in a black-and-white photograph two colors, for example red and green, will appear as the same shade of gray; hence any border between them will vanish. Similarly, at some relative red-to-green brightness level, the red and green will appear identical to the magno system. The red and green are then called equiluminant. A border between two equiluminant colors has color contrast but no luminance contrast. The exact brightness ratio at which a border between two colors becomes invisible to the magno system varies according to the person, just as different



SHADING can create a strong sense of depth, as is demonstrated in this computer-generated image (*left*). When the shading is produced by color contrast instead of luminance contrast (*right*), the sense of three-dimensional shape is reduced and the image becomes almost impossible to see. This effect suggests that depth perception is processed by the color-blind magno system.

black-and-white films may vary in their sensitivity to certain colors.

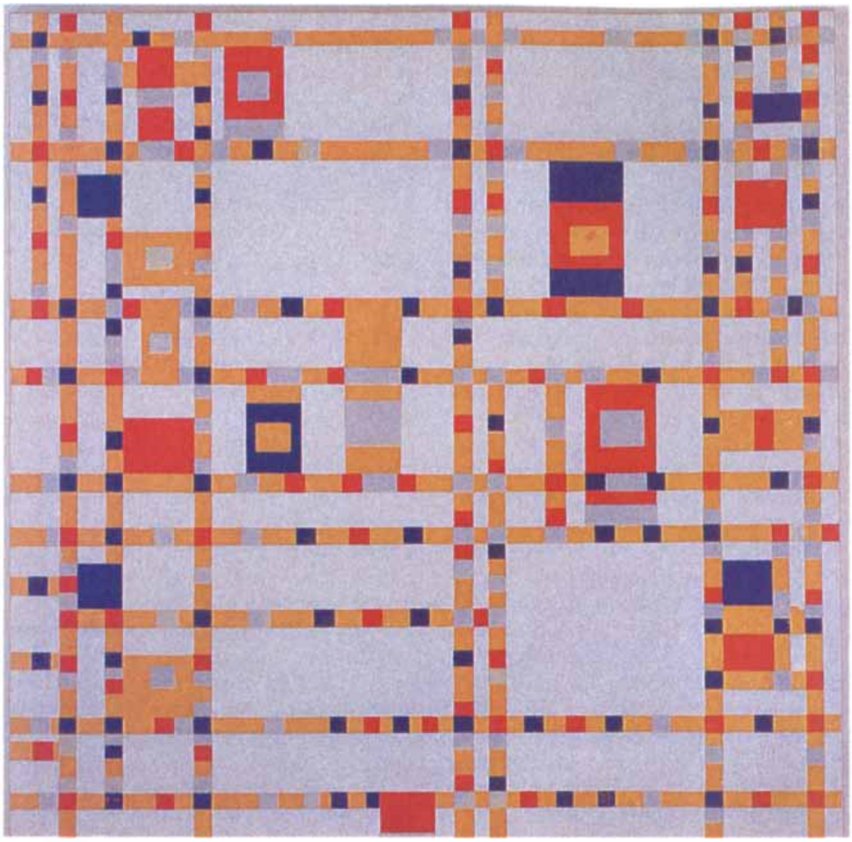
The fact that the visual system is divided into three parts, each with quite different characteristics, led Hubel and me to speculate that such differences might be reflected in human perception. We set out to test this by examining the four properties that differentiate the magno and parvo divisions at their earliest stages: color, contrast sensitivity, acuity and speed. We wanted to know, for example, if movement and depth perception are color- or contrast-sensitive and whether or not they have low spatial and high temporal resolution. A review of psychological literature spanning more than a century of research provided answers that were wonderfully consistent with the functional segregation suggested by the anatomical and physiological data.

One of our questions had already been answered by Patrick Cavanagh, Christopher Tyler and Olga E. Favreau of the University of Montreal, who had demonstrated that movement perception is color-blind. They found if they created red and green moving stripes on a television screen and then adjusted them so that they were equiluminant, either the perceived speed of movement was greatly reduced or the stripes appeared to completely stop moving on the screen. Further tests revealed that movement perception also has low acuity, high-contrast sensitivity and a fast response time. All these observations are consistent with the idea that motion perception is mainly a function of the magno system and not the parvo system.

Cary Lu and Derek H. Fender of Caltech performed a series of related experiments in which they tested the color sensitivity of stereopsis. Because of the distance between the eyes, a three-dimensional scene produces slightly different images on the two retinas and the visual system interprets these differences as distance. Lu and Fender found that people cannot see depth in a stereoscopic image when the stereoscopic stimuli are equiluminant. This suggests that stereopsis, like movement perception, is color-blind. Other aspects of stereopsis also seem to conform to the distinguishing characteristics of the magno system.

That finding led us to conclude that these two functions are probably carried almost exclusively by the magno system. We wondered whether other aspects of vision would show similar constellations of characteristic properties. Stereopsis is the depth cue studied most often by visual physiologists, probably because it can be quantified easily. But it is not the only depth cue and is not essential for depth perception, as a person can quickly verify by closing one eye. If, in looking at photographs or paintings, you experience any impression of depth, you must necessarily be ignoring or overriding the information from stereopsis that the picture is in fact flat. This is why, as Leonardo da Vinci pointed out, a person can increase the impression of depth in a painting by viewing it with only one eye.

We wanted to know whether other nonstereoscopic depth cues might also be color-blind. Included in this category are such cues as perspective and the relative size of objects, the relative movement of objects (when you move your head sideways, objects closer to you move farther across the retina than far objects do) and shading and gradations in texture. We found that all these cues are also color-blind, which leads us to suspect that most depth information is carried by the magno system.



OP ART often appears jumpy, as if the colors are moving or unstable, a phenomenon related to luminance. That is seen here in the Piet Mondrian painting *Broadway Boogie Woogie*, where the yellow stripes have low luminance contrast against the off-white background. The impression of movement is induced by the lessened ability of the brain to assign a stable position to the yellow stripes, and so they seem to jump around.

Shading is an important cue for the perception of depth and shape, which is why it is so much harder to see bumps on a ski slope on a heavily overcast day. Under most natural light conditions, shadows differ in their degree of brightness rather than in their color. Therefore the part of the visual system that recognizes shapes according to their shading does not have to be color-blind, but it certainly has no need to carry color information. Cavanagh and his colleague Yvan LeClerc demonstrated this in a series of experiments in which they were able to show that a shadow can be any hue; it needs only to be darker than the rest of the surface to convey a sense of depth.

We wondered why the magno system should respond both to movement and to the many cues used to judge distance and spatial relations. Why should this particular assortment of functions be carried by only one system? One possible answer comes from the turn-of-the-century school of thought known as Gestalt psychology. Most images contain a rich assortment of visual elements: edges of many orientations and surfaces, colors and textures that are homogeneous. Perception requires that these various elements be organized in such a way that related ones (those belonging to a single object) are grouped together.

The Gestalt psychologists suggested that this is possible because the brain uses certain visual properties to group the parts of an image together and also to separate images from one another and from their background. These properties include direction and velocity of motion (elements that move together probably belong to the same object); collinearity (a house is not perceptually split in two when a telephone wire crosses in front of it); depth (two borders may be contiguous but do not appear to belong to the same object unless they are at the same distance from the viewer), and lightness and texture (different parts of the same object usually have the same surface properties). The fact that these functions all fail at equiluminance suggested to us that the ability to link parts of a scene together, to discriminate figure from ground and to perceive the correct spatial relationship of objects might all be carried by the magno system. We surmised that the magno system must combine the visual properties of an object in a way that enables it to perceive the image as a whole, thereby freeing the parvo system to see details.

What implications do these ideas have for art and design? Although the neurobiological explanations for many of the phenomena I have described have only recently been clarified, many artists and designers seem to be empirically aware of the underlying principles. Understanding how the brain processes visual information may make it even easier for artists and designers to maximize particular desired effects.

Some of the peculiar effects of op art, for example, probably arise from color combinations that are strong activators of the parvo system but are weak stimuli for the magno system. An object that is equiluminant with its background looks vibrant and unstable. The reason is that the parvo system can signal the object's shape but the magno system cannot see its borders and therefore cannot signal either the movement or the position of the object. Hence it seems to jump around, drift or vibrate on the canvas.

COLOR BLENDING occurs when patterns or designs are too fine to be seen by the lower-resolution color system. Pointillist paintings, such as this one by Georges Seurat (*left*), create this effect when they are made of dots large enough to be seen by the form system but too fine for the color system. When the painting is viewed close up (*right*), however, the dots can be seen by both the color and the form systems and the colors no longer blend.

Advertisements often contain key words in a color that is equiluminant with the background. Although it is more difficult to read the text under those conditions, one possible explanation for the choice of color is that the weird, jumpy appearance created by equiluminance draws the attention of the viewer. Whether accidental, empirical or conscious, the result is that the reader must spend more time reading those words in order to decipher them.

The same principles apply to fashion. The overall impression of a garment's shape can be influenced by lines in the material or trim, but the lines must have luminance contrast to be effective; the higher the luminance contrast, the greater the effect. Conversely, patterns will not interfere with the line of a garment if they are formed by equiluminant colors. Horizontal stripes, for example, will not make the wearer appear shorter and wider if they are close to equiluminant or are very narrow. The combination of a Kelly green shirt with royal blue pants produces a jazzy, vibrant, horizontal border that will not contribute much to the overall impression of shape or draw the eye. If, however, the same pants are combined with a shirt of similar hue but a different shade, such as sky blue or navy, then the eye is drawn to the border and the overall effect is less vertical.

The color-selective blob system has a severalfold lower acuity than the interblob system. This lower resolution explains a phenomenon that has been known for at least a century: two colors can have entirely opposite effects on each other depending on their spatial arrangement. When two different colors are juxtaposed, they normally oppose each other, so that each appears less like the other (that is, each tends toward the complement of the other). On the other hand, when the two colors are interdigitated in a fine pattern, there is an opposite effect: they come to look more like each other, that is, they blend or "bleed." We believe that blending occurs when the pattern is too fine for the color system to resolve.

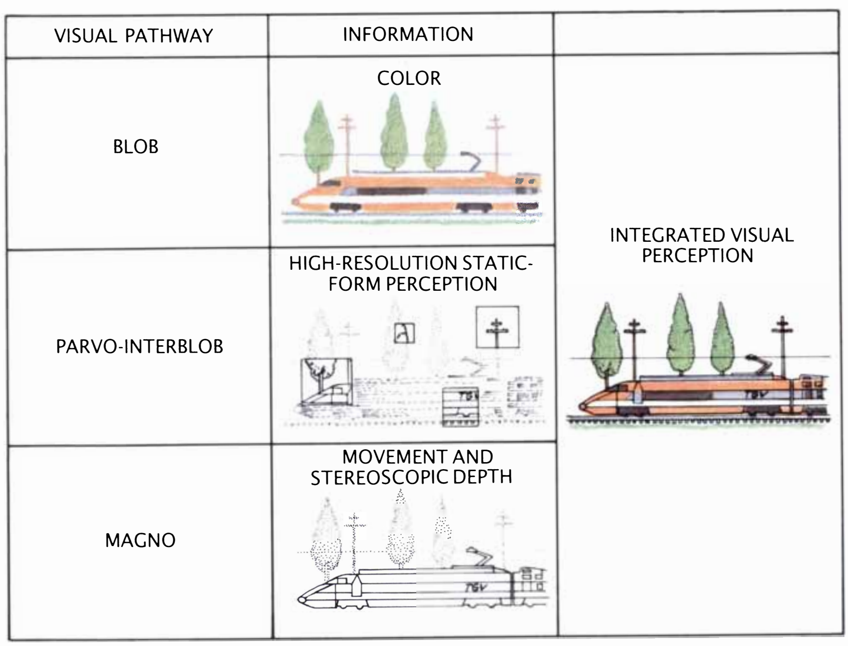
A similar effect is seen in magazine illustrations when the microscopic dots that form the image blend together, because humans cannot discriminate the small dots with either the form or the color system. The phenomenon only becomes noticeable when the pattern, like the brushstrokes in an impressionist painting or the dots in a pointillist painting, is too small to be resolved by the color system but is large enough to be resolved by the form system. Then an observer can see the individual brushstrokes or the dots, but the colors blend anyway. There is, then, a relatively narrow range of pattern sizes over which a person can discriminate the pattern but still not resolve discrete blocks of color. That range is defined by the difference in resolution between the color system and the form system. A similar blending of colors, without a blending of their pattern, can often be seen in finely patterned fabrics such as tweeds or pinstripes.

The television industry takes advantage of the low resolution of the color system by transmitting the color part of an image at a lower resolution than the black-and-white part, thereby reducing the total amount of information that needs to be transmitted. The lower acuity of the color system also explains why many artists working with watercolors or pastels can apply color quite loosely to an outlined area and yet produce the same visual effect as if the color were applied precisely within those areas.

Knowledge about the segregation of functions in the visual system may be useful to artists, advertisers and designers; it may also be useful in designing surveillance systems, in devising ways to enhance the visibility of blurry, low-contrast or camouflaged images and in designing video systems for robotics or automatic navigation. The sensitivity of the magno system to color contrast, movement and stereopsis suggests that hard-to-see objects can be made more visible by introducing movement (by moving the object or the observer) or stereopsis (by simultaneously viewing two images of the same scene taken from different positions). Such an approach might enable radiologists to interpret fuzzy images, such as X rays, with greater accuracy.

Until recently only a few aspects of perceptual psychology and aesthetics could be correlated with what was known about how the brain processes information. Now that situation is changing, and vision research is at an exciting time in its history. Art, psychology and neurobiology are each beginning to contribute information to one another and in the process are becoming mutually enriching disciplines.

IMPRECISE POSITION of color in many watercolor paintings, as in this reproduction of Pablo Picasso's *Mother and Child,* does not interfere with the assignment of colors to objects in the image. The colors appear to conform to the outlined shapes more closely than they really do, because the color system has a three- to fourfold lower acuity than the form system. The technique is most effective with watercolors and pastels because the colors are pale and therefore do not create a strong contrast with the background.



VISION is a three-part system. Color is handled by the blob pathway, static-form perception by the parvo-interblob pathway, and movement and depth by the magno pathway. The end result is integrated so that one sees a unified, three-dimensional world.

The Transformer

Just a century ago this inconspicuous device made the distribution of electric power a practical endeavor. Many elements of modern life depend on it, yet it remains one of technology's unsung heroes

by John W. Coltman

The technological revolution that has shaped civilization for the past 100 years sprang from fundamental advances in communications, transportation and electric power. The crowning achievements of inventors in communications and transportation—the telephone, television, the automobile, the airplane—are by now familiar fixtures of everyday life. In contrast, the invention that ensured the ubiquity of electric power goes largely unrecognized by those whose lives are touched by it. It is a device that does not move, is almost totally silent and is typically hidden in underground vaults or stowed behind screens.

That device is the transformer, an ingenious instrument developed late in the 19th century. The transformer is an essential component of modern electric power systems. Simply put, it can convert electricity with a low current and a high voltage into electricity with a high current and a low voltage (and vice versa) with almost no loss of energy. The conversion is important because electric power is transmitted most efficiently at high voltages but is best generated and used at low voltages. Were it not for transformers, the distance separating generators from consumers would have to be minimized; many households and industries would require their own power stations, and electricity would be a much less practical form of energy.

In addition to its role in electric power systems, the transformer is an integral component of many things that run on electricity. Desk lamps, battery chargers, toy trains and television sets all rely on transformers to cut or boost voltage. In its multiplicity of applications the transformer can range from tiny assemblies the size of a pea to behemoths weighing 500 tons or more. This article will focus on the transformers in power systems, but the principles that govern the function of electrical transformers are the same regardless of size or application.

The English physicist Michael Faraday discovered the basic action of the transformer during his pioneering investigations of electricity in 1831. Some 50 years later the advent of a practical transformer, containing all the essential elements of the modern instrument, revolutionized the infant electric-lighting industry. By the turn of the century alternating-current power systems had been universally adopted and the transformer had assumed a key role in electrical transmission and distribution.

Yet the transformer's tale does not end in 1900. Today's transformers can handle 500 times the power and 15 times the voltage of their turn-of-the-century ancestors; the weight per unit of power has dropped by a factor of 10 and efficiency typically exceeds 99 percent. These advances reflect the marriage of theoretical inquiry and engineering that first elucidated and then exploited the phenomena governing transformer action.

Faraday's investigations were inspired by the Danish physicist Hans Christian Oersted, who had shown in 1820 that an electric current flowing through a conducting material creates a magnetic field around the conductor. At the time Oersted's discovery was considered remarkable, since electricity and magnetism were thought to be separate and unrelated forces. If an electric current could generate a magnetic field, it seemed likely that a magnetic field could give rise to an electric current.

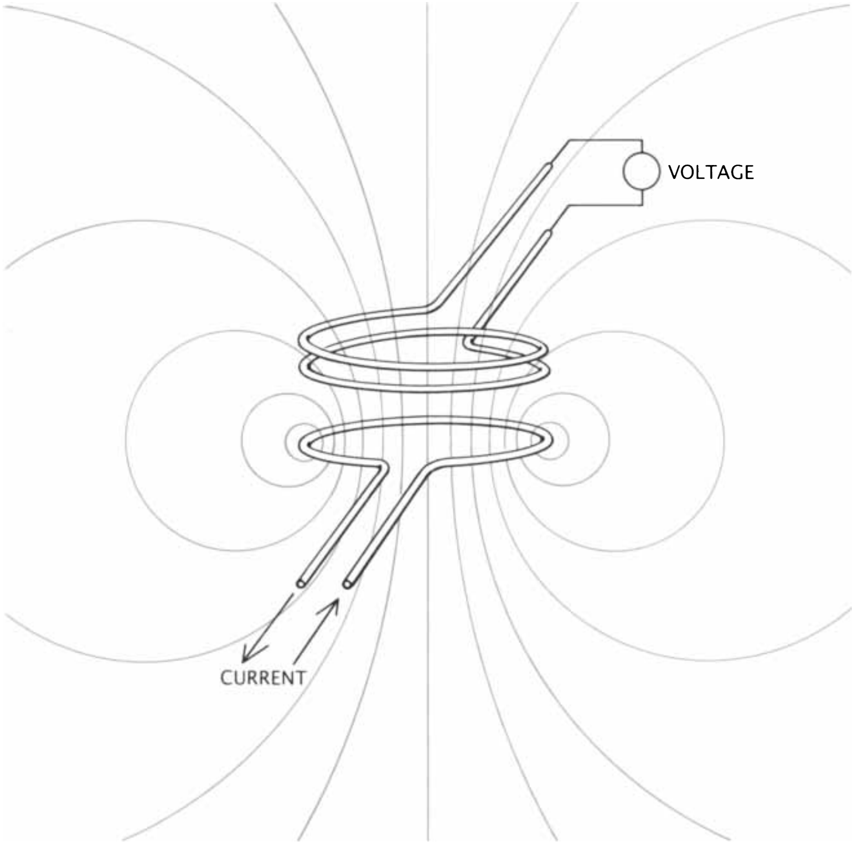
In 1831 Faraday demonstrated that in order for a magnetic field to induce a current in a conductor the field must be changing. Faraday caused the strength of the field to fluctuate by making and breaking the electric circuit generating the field; the same effect can be achieved with a current whose direction alternates in time. This fascinating interaction of electricity and magnetism came to be known as electromagnetic induction.

Induction can best be understood in terms of lines of force, a convention Faraday introduced in order to describe the direction and strength of a magnetic field. The lines of force for the magnetic field generated by a current in a loop of wire are shown in the illustration on page 88. If a second, independent loop of wire is immersed in a changing magnetic field, a voltage will be induced in the loop that is proportional to the time rate of change of the number of force lines enclosed by the loop. If the loop has two turns, such induction occurs in each turn, and twice the voltage results; if a loop has three turns, three times the voltage results, and so on. (Voltage can be thought of as the pressure that drives a charge; current is the rate of flow of charge. The product of these—voltage in volts multiplied by current in amperes—equals the electric power in watts.)

In a transformer the loop of wire that is fed the current and that generates the magnetic field is called the primary. The loop that intercepts the field is called the secondary. Induction between the primary and the secondary is mutual; that is, a current flowing in the secondary will induce a voltage in the primary in the same way as the primary induces a voltage in the secondary. Furthermore, since the primary loop encloses its own lines of force, it can induce a voltage in itself; this process is known as self-induction, and it takes place in the secondary as well.

The concurrent phenomena of mutual induction between the coils and

TRANSFORMER at a New York City power station converts electricity from low voltages into the high voltages required for transmission. Radiators (*left*) cool the transformer, and cylindrical bushings carry power lines to and from it. Other remote transformers cut the voltage to levels appropriate for distribution and consumption (*see illustration on pages 94 and 95*).



LINES OF FORCE describe the magnetic field emanating from a coil of wire (the primary) carrying a current. A second coil (the secondary) placed in the field intercepts the lines (*color*); if the magnetic field is fluctuating, as it is when the primary coil is fed alternating current, it will induce a voltage in the secondary coil. This phenomenon, which is known as electromagnetic induction, is the foundation of transformer action.

self-induction in each coil are at the heart of transformer action. In order for a power transformer to do its job effectively, the coils must be almost perfectly coupled and each must have high self-induction. That is, almost all the lines of force enclosed by the primary must also be enclosed by the secondary, and the number of force lines produced by a given rate of change of current must be high. Both conditions could be met by wrapping the primary and secondary coils around an iron core as Faraday did in his first experiments. Iron increases the number of lines of force generated by a factor of about 10,000, a property known as permeability. It also constrains the lines so that the primary and secondary coils can be spatially separated and still be closely coupled magnetically.

In the ideal transformer all the lines of force go through all the turns in both coils, and since a changing magnetic field produces the same voltage in each turn of a coil, the total voltage induced in a coil is proportional to the total number of turns in that coil. If no energy is lost in the transformer, the power available in the secondary must be equal to the power fed into the primary; in other words, the product of the secondary voltage and the secondary current equals the product of the primary voltage and the primary current. Thus the two currents must be inversely proportional to the two voltages, and therefore inversely proportional to the turns ratio between the two coils. (The expressions of power are true only if the currents and voltages are in phase; the condition of high self-induction ensures that out-of-phase currents will be negligible.)

Such an ideal transformer provides the electrical engineer with a tool quite like the lever in mechanics, but instead of converting force and motion, the transformer deals in voltage and current. Instead of lever-arm length, the turns ratio is the operative feature of the instrument. Of course, the ideal transformer has not yet been devised, but it has been closely approached in practice. Iron cores are essential components of all modern power transformers, and copper, because of its low electrical resistance, was and still is the material of choice for the coils.

The simple relations in an ideal transformer, ABC's for the present-day electrical engineer, were by no means clear to the early experimenters. The arrangements they worked with were far from ideal and the combined phenomena of self- and mutual induction, with poorly coupled coils and imperfect iron, gave rise to much complex and mysterious behavior.

Faraday left his ruminations without carrying them much further, certain that other inventors would pick up where he left off. Actually for several decades there simply were no general applications for transformer-like devices. Initial experiments with "inductors" having a single wire wrapped around an iron core were marked by wonder at their ability to generate sparks when the current supplied to the coil was interrupted. Among the eminent scholars who explored this phenomenon was the American Joseph Henry, first secretary and director of the Smithsonian Institution, after whom the unit of induction is named.

During this period of experimentation it became apparent that currents circulating in solid metal cores were wasting energy. In order to minimize these so-called eddy currents, cores were constructed that were nonconducting in the direction perpendicular to the magnetic lines of force in the transformer. This was accomplished by making the core out of a straight bundle of iron wire.

All the work of this period was carried out with batteries as sources, the primary circuit being closed and opened to produce the necessary changing current. In the 1860's the introduction of the dynamo—an electric generator also based on Faraday's insights—made alternating current generally available. The first person to connect a transformer to an a.c. source was Sir William Grove, who needed high-voltage power for his laboratory work. In the absence of an obvious commercial application, however, the significance of the arrangement was overlooked, and it remained obscure until Thomas Edison began to promote the idea of an electric-lighting system in the 1880's.

When Edison launched his scheme, light bulbs equipped with platinum filaments were already available; arc lighting, in which current "arcs" be-

tween two electrodes rather than passing through a filament, was also in use. Both kinds of lamps worked well, but their electrical characteristics placed some constraints on the way they could be wired together. In particular, the lamps had to be connected in series in one continuous circuit, just like old-fashioned Christmas-tree lights. Consequently all the lights in the system had to be turned on or off simultaneously.

Although such an arrangement was acceptable for applications such as street lighting, the inability to turn individual lamps on and off at will, and the very high voltages present in the system when a large number of lamps were joined in series, militated against series electric lighting in houses and small installations. On the other hand, parallel systems, in which each lamp operates on its own "subcircuit," required impractically large copper wires to supply the low-resistance, high-current lamps of the day. Edison's major accomplishment was the introduction of a carbon-filament lamp that, because of its high resistance, made parallel connection feasible. In New York City, with these lamps and a direct-current power generator, Edison opened the first commercial lighting plant in 1882.

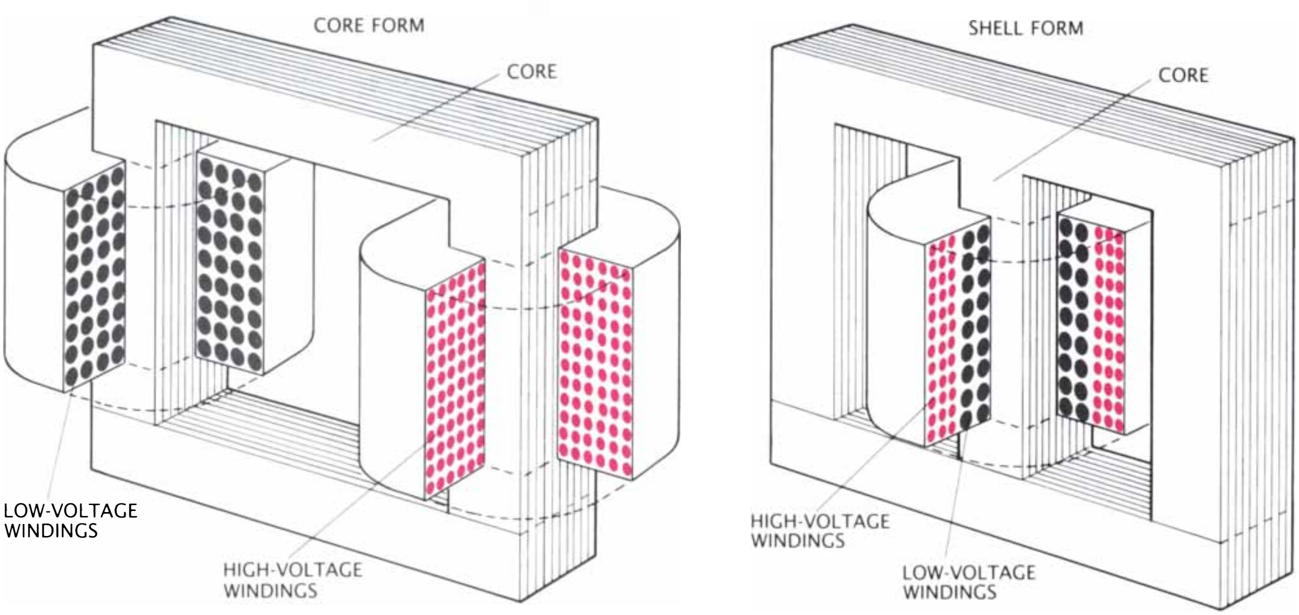
At about the same time transformers were first incorporated in an electric-lighting system in England. Lucien H. Gaulard and John D. Gibbs—a French inventor and an English promoter—used a form of transformer to add incandescent lamps to an a.c. arc-lighting system. Because the arc lamps were connected in series, with a fixed current running through the line, the primaries of their transformers were necessarily in series with the arc lamps. Gaulard and Gibbs were granted a patent for the device, which they called the secondary generator, in 1882, and they demonstrated their system in England in 1883 and in Italy in 1884. The secondary generator was not a very practical piece of equipment; it saw little actual use, but it stimulated thought among other inventors.

Among those who became interested in Gaulard and Gibbs's work were three Hungarian engineers from Ganz and Company in Budapest. They saw the demonstration in Italy and recognized the disadvantages of series connection. When they returned to Budapest, Max Déri, Otto T. Bláthy and Karl Zipernowski built several transformers for parallel connection to a generator. The engineers designed two types of transformers with closed iron cores that were much better adapted to parallel connection than the open-ended bundles: one in which the conductors were wound around a toroidal, or doughnut-shaped, core and one in which the wires of the iron core were wound around a toroidal bundle of conductors.

In May of 1885, at the Hungarian National Exhibition in Budapest, Déri, Bláthy and Zipernowski demonstrated what is generally considered to be the prototype of today's lighting systems. Their system included 75 transformers in parallel connection, powering 1,067 incandescent Edison lamps from an a.c. generator supplying 1,350 volts. The transformers had toroidal iron cores with the conductors wound laboriously around them. Although they were expensive to build, they were efficient enough to feasibly carry out the function for which they were designed: to operate low-voltage lamps from a high-voltage distribution system.

An American named George Westinghouse was also impressed by the Gaulard and Gibbs demonstration in Italy. In the 1880's Westinghouse was already an established inventor and industrialist and was working on the distribution of natural gas for illumination. At the time of Edison's success he became interested in electric power, but he was wary of its applicability. His skepticism was well founded: in a parallel system increased load demands increased current, and a "load" the size of a city would require huge amounts of current. But transmission of high-current power is inefficient; it would therefore be necessary to send the power over enormous copper wires or to build generating plants quite close to their loads, scattering many small plants throughout a large city.

Efficient transmission of high-volt-

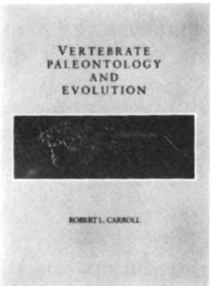


TWO TRANSFORMER DESIGNS illustrate different approaches to core structure and wiring. Both cores are made from stacked laminations stamped out of iron sheets. In the design at the left, called core form, the primary encloses one arm of the core and the secondary the other. The shell-form core at the right is made up of E-shaped stampings with the primary and secondary coils nested together on the middle bar. In three-phase transformers the coils are nested on all three bars (*see illustration on page 93*).

Save 20% on these new titles from W. H. Freeman and Company

The book publishing arm of *Scientific American*

Vertebrate Paleontology and Evolution



Robert Carroll,
REDPATH MUSEUM, McGILL UNIVERSITY

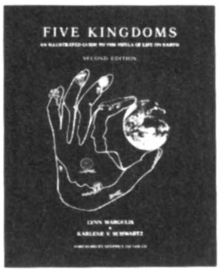
"A real tour de force."
—Léo Laporte,
University of California, Santa Cruz

Vertebrate Paleontology and Evolution is the first detailed, authoritative review of the fossil record in twenty years. Incorporating the latest discoveries and newest ideas about evolutionary processes, it provides complete, current data on all fossil groups and examines each in terms of classification, skeletal and functional anatomy, way of life, distribution in time and space, and interrelationships with other groups. Featuring more than 1700 illustrations and a massive bibliography containing nearly 1500 citations, this is an indispensable reference to the fascinating history of fossil vertebrates through 600 million years of evolution.

698 pages, 1779 illustrations
Regular Price $52.95 **YOUR PRICE $42.59**

Five Kingdoms An Illustrated Guide to the Phyla of Life on Earth

Second Edition

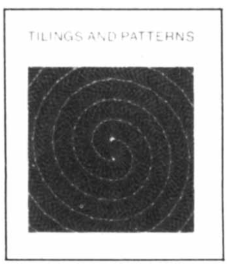


Lynn Margulis, BOSTON UNIVERSITY
Karlene V. Schwartz,
UNIVERSITY OF MASSACHUSETTS, BOSTON
Foreword by Stephen Jay Gould

This profusely illustrated guide classifies all life on Earth into the 92 phyla that make up nature's five great kindgoms: Prokaryotae, Protoctista, Fungi, Animalia, and Plantae. The Second Edition is completely up-to-date and includes three newly-recognized phyla—including the first new animal phylum to be acknowledged in 30 years—and incorporates new ideas in both protoctistan and bacterial relationships. The Second Edition includes bibliographies, an appendix listing the phylum and common name of some 1000 genera, a glossary of technical terms, and an extensive index.

416 pages, 421 illustrations
Regular Price $35.95 **YOUR PRICE $28.79**

Tilings and Patterns

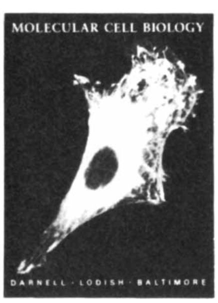


Branko Grünbaum,
UNIVERSITY OF WASHINGTON
Geoffrey C. Shephard,
UNIVERSITY OF EAST ANGLIA, ENGLAND

This is the first comprehensive and systematic approach to the classification and enumeration of tilings. "I am enormously impressed.... Will be the definitive reference on tiling theory for many decades. Not only does the book bring together older results that have not been brought together before, but it contains a wealth of new material.... I know of no comparable book." *—Martin Gardner*

700 pages, 1,406 illustrations
Regular Price $59.95 **YOUR PRICE $47.99**

Molecular Cell Biology



James Darnell, ROCKEFELLER UNIVERSITY
Harvey Lodish, INSTITUTE FOR BIOMEDICAL RESEARCH and M. I. T.
David Baltimore, DIRECTOR, WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH, and M. I. T.

Darnell-Lodish-Baltimore's *Molecular Cell Biology* is the new standard in the field, offering the most authoritative and up-to-date treatment of cell biology and molecular genetics available today. "A didactically effective and intellectually elegant work.... An exceptionally versatile book [that] will meet the needs of both students and more advanced investigators and teachers...."
—Thoru Pederson, in *Cell* (12/5/86)

1,187 pages, 1,248 illustrations
Regular Price $42.95 **YOUR PRICE $34.59**

Please send me:

Quan	ISBN	Title	Special Low Price	Amount
	1822-7	Vertebrate Paleontology	$42.59	
	1885-5	Five Kingdoms	$28.79	
	1193-1	Tilings and Patterns	$47.99	
	1448-5	Molecular Cell Biology	$34.59	
			TOTAL	
			Add $1.50 for shipping and handling, plus $1.00 for each additional book	
			Sales tax (NY, NJ, CA, and UT)	
			FINAL TOTAL	

☐ I enclose a check made payable to W. H. Freeman and Company

☐ Charge my:

☐ MasterCard ☐ VISA Account #________

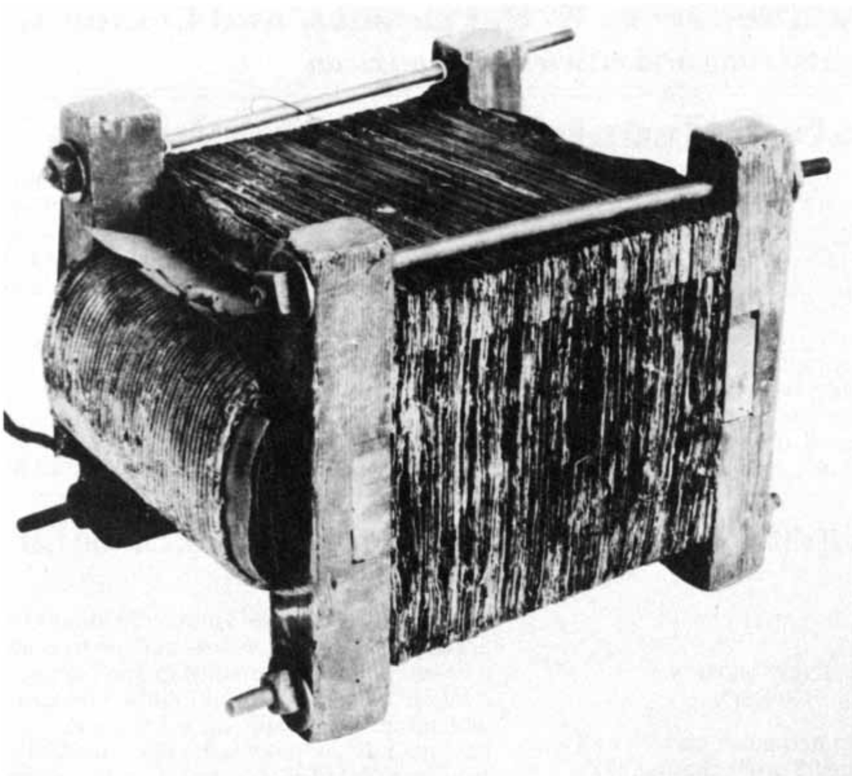
Expiration date____/____ Signature________
(Credit card orders must be signed)

Name________

Address________

City / State / Zip

Mail to: **W. H. Freeman and Company**
4419 West 1980 South, Salt Lake City, Utah 84104



STANLEY TRANSFORMER from the first a.c. power station in Great Barrington, Mass., dates from 1885. The transformer is about a foot long; copper windings wrapped with cotton protrude between wood endpieces at the left. The middle arm of E-shaped iron laminations was slid into the prewound coil in alternating directions. The ends of the other two arms are visible as dense regions at the top and bottom of the laminations.

age power, on the other hand, was possible with relatively small wires, and many people were looking for ways to transmit electric energy at voltages higher than those required at the point of application. In 1884 Westinghouse hired a young engineer, William Stanley, who already had some ideas about solving the problem with transformers. When he heard about Gaulard and Gibbs's work, he encouraged Westinghouse to take an option on the transformer patents. Stanley was convinced of the superiority of parallel connection, and by the early summer of 1885 he had designed some closed-core transformers.

Soon afterward health problems made it prudent for Stanley to set up a laboratory away from the smoky Pittsburgh atmosphere of the time. With Westinghouse's approval he moved to Great Barrington in Massachusetts and continued his work on transformers. In the meantime Westinghouse, who was not entirely convinced of the wisdom of parallel connection, explored various combinations of the Gaulard and Gibbs secondary generators with another pioneer in electrical engineering, Oliver B. Shallenberger.

By December, 1885, Stanley had made enough progress to win Westinghouse over. With the help of Shallenberger and another brilliant engineer, Albert Schmid, Westinghouse set about modifying Stanley's transformer so that it (unlike the Hungarian toroidal type) could be manufactured easily and cheaply. The core was made of thin sheets of iron cut at first in the form of the letter H. Coils of insulated copper wire were wound around the crossbar of the H and then the ends of the H were closed with separate strips of iron. Stanley suggested making the iron stampings in the form of an E so that the center prongs could be slid into a prewound coil. The E-shaped stampings were inserted in alternating directions, and straight pieces of iron were laid across the ends of the arms to complete the magnetic circuit. This construction is still common today.

The Westinghouse Electric Company was chartered in January of 1886. Over the next few months Westinghouse and his associates patented the process for inserting stacked iron plates into prewound coils, the provisions for cooling and insulating the transformer by immersion in oil and the packaging of the assembly in a hermetically sealed container. Stanley constructed and installed several transformers in Great Barrington and wired the system for 500-volt distribution from the laboratory to the town center, a distance of almost a mile. To demonstrate the possibility of efficient transmission over longer distances, he also used transformers to step up the electric power to 3,000 volts and then to cut it down to 500 volts before sending it out on the town line. On March 16, 1886, Stanley's plant went into service. It was a great success, and Westinghouse proceeded to establish facilities for the manufacture and sale of equipment for distributing alternating-current electric power.

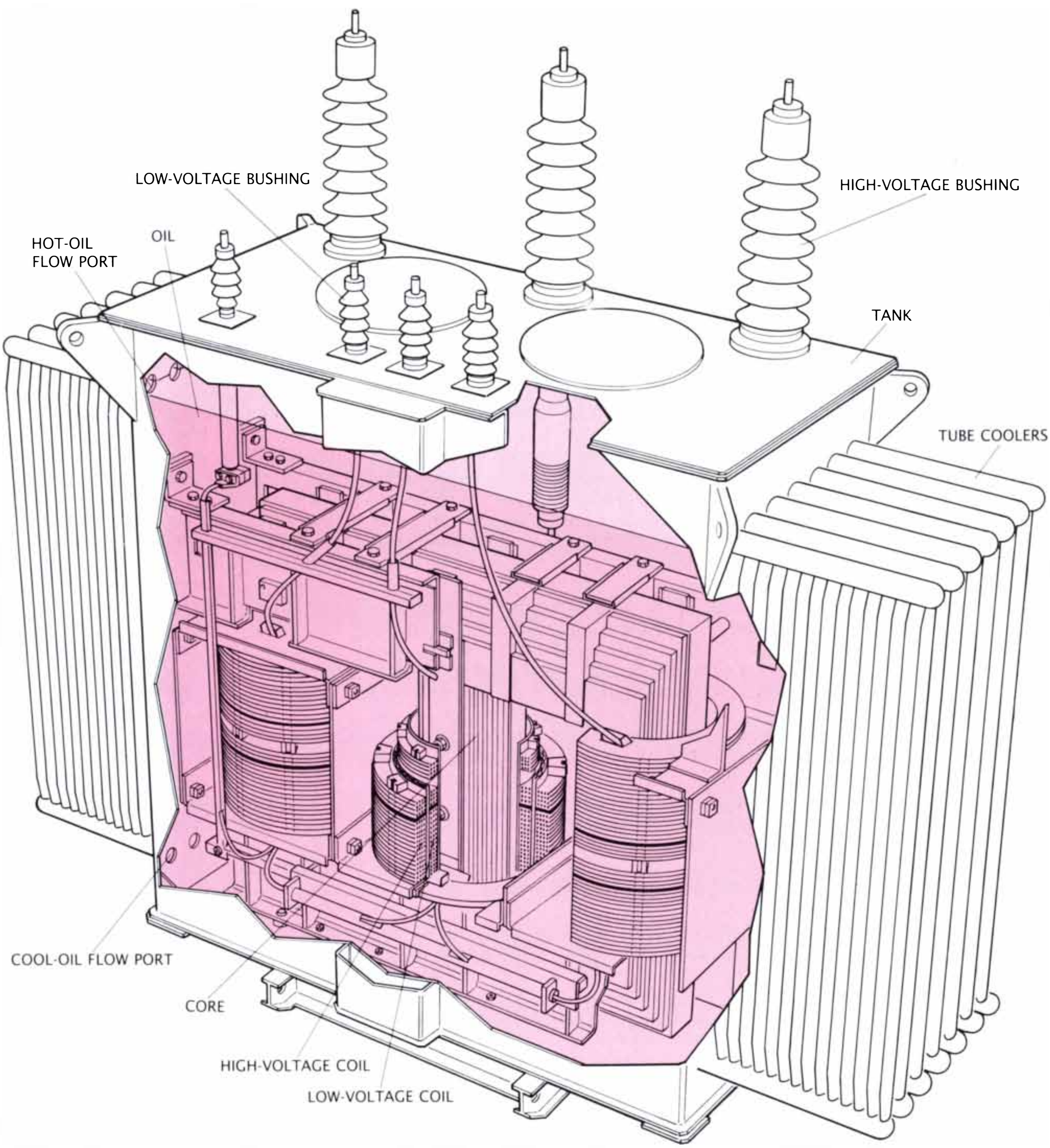
Edison and his associates fought the alternating-current system in both the courts and the press, but theirs was a losing battle. The polyphase motor invented by Nikola Tesla provided an efficient way to utilize alternating current, and Shallenberger's invention of the a.c. watt-hour meter made possible accurate billing of customers for energy consumption. These two inventions, together with the low cost of transmitting alternating current, gave the a.c. system a flexibility and convenience that soon relegated d.c. systems to a few specialized applications.

The next decade saw the rapid growth of a.c. electric power systems, marked by achievements such as the lighting of the 1893 world's fair in Chicago and the installation at Niagara Falls of huge 5,000-horsepower hydroelectric generators, the first two of which went into service in 1895. Along with the staggering growth of electrical generating power came great increases in the size of transformers. In 1895 a furnace at the Carborundum Company in Niagara Falls employed a transformer rated at 750 kva (kilovolt-amperes, a designation of power-handling capacity akin to kilowatts). Five years later some transformers were rated at 2,000 kva and operated at 50,000 volts.

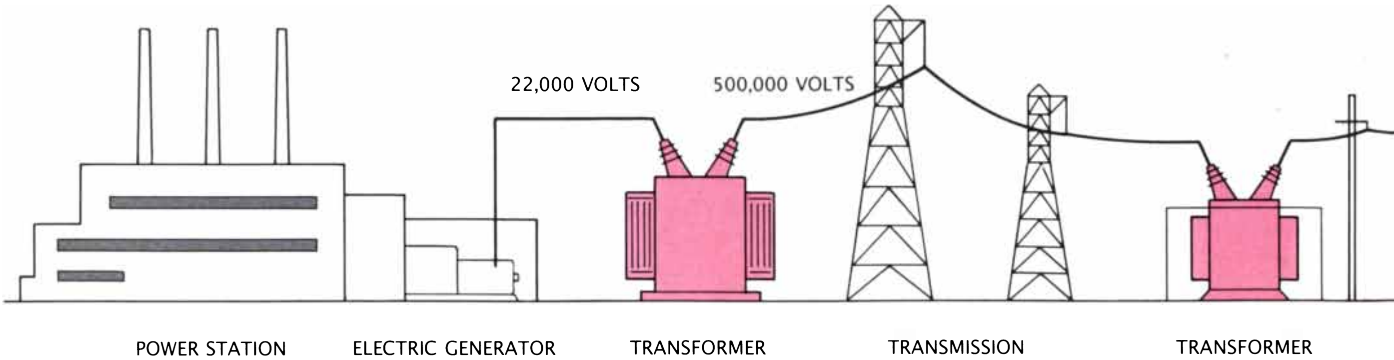
It could be argued that the transformer built at the turn of the century was already a mature product: the essential features of the device remain unchanged to this day. In fact, however, the transformer continued to evolve. Although it is still a cooled and insulated assemblage of iron sheets and copper coils, the improvement in transformer performance since 1900 has been quite remarkable. Modern transformers can operate at 765 kilovolts and handle more than a million kva, and they have lifetimes of from 25 to 40 years.

These improvements give testimony to the efficacy of the industrial-research process, a process whose rapid growth was closely associated with the rise of the electric power industry. The practitioners of industrial research, driven by a competitive system that rewards maximum performance at minimum cost, seek an understanding of natural phenomena in order to develop new products and processes and to improve old ones. Competition provides the impetus for eliminating the limitations imposed by materials while at the same time giving rise to better designs and fabrication methods that take advantage of improved materials and fresh insight.

The parameters that characterize the ideal transformer depend to a



TYPICAL MODERN TRANSFORMER is submerged in oil for insulation and cooling and is sealed in an airtight tank. Low- and high-voltage power lines lead to and from the coils through ceramic bushings. Inside the transformer, coils and core are packed close together to minimize electrical losses and material costs. The oil coolant circulates by convection through external radiators. In large transformers cooling is expedited by attaching fans to the radiators and circulating the oil with pumps.



PATH OF ELECTRIC POWER begins at the power plant, where electricity is generated at about 22,000 volts and is stepped up for transmission on high-tension lines. Step-down transformers reduce the voltage to levels appropriate for local distribution

large extent on the properties of the core, and it is in the core that the most significant advances have been made. The properties that are important in a core material are permeability, saturation, resistivity and hysteresis loss. Permeability, as mentioned above, refers to the number of lines of force a material produces in response to a given magnetizing influence; saturation designates the point at which the material's ability to amplify an external magnetizing force reaches a plateau. These two properties define the power-handling capability of the core. Electrical resistivity is desirable in the core because it minimizes energy losses due to eddy currents.

In contrast, hysteresis, the "memory effect" in magnetic materials, undermines the efficiency of transformer action. Because of the interactions among groups of magnetized atoms, the effects of magnetization tend to "stick" in a material, so that if the magnetizing force is lowered temporarily the material does not respond right away. In a transformer the lag translates into energy wasted during every cycle of alternating current. Throughout the history of core development the goal of the engineer has been to increase permeability, saturation and resistivity while decreasing hysteresis losses.

One of the more important tools in this quest is the B-H curve, which graphically describes the relation in a given magnetized material among permeability, saturation and hysteresis. It is a plot of the number of lines of force induced in a material (B) as a function of a varying magnetizing force (H). Shaped like a broad S tapered at each end, the curve is traced out on each of the cycles of the alternating driving current. Its slope corresponds to the permeability, the point at which it levels out (the top of the S) is the saturation value and the area under the curve (the area of the S) corresponds to the hysteresis.

The goal of the scientist has been to find out how these properties are related to the physical constitution of iron. Each property depends on cooperative interactions among the atoms in elementary magnets, which are affected by the crystal structure of iron and the presence of other elements and imperfections. The study of these complicated interactions is called domain theory; the insight it provides guides experimenters in their search for better transformer materials.

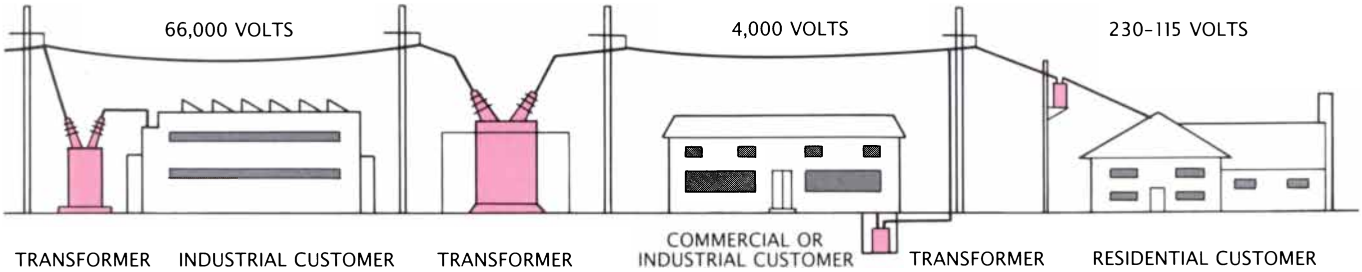
The thin wrought-iron sheets of which cores were made in the first Stanley-Westinghouse transformers had substantial hysteresis losses that were gradually lessened by selecting iron from particular manufacturing sources, so that by 1900 the losses had been cut in half. Aging of the material was also a problem: as the transformer grew older, hysteresis losses became progressively worse.

In the early 1900's an English metallurgist, Robert A. Hadfield, was engaged in a long series of experiments aimed at determining how the properties of iron were affected by the addition of other elements. In a number of papers Hadfield and his colleagues revealed the potential of silicon iron as a core material. Adding silicon to iron reduced hysteresis losses, increased permeability, virtually eliminated aging and increased the electrical resistivity of the metal. Silicon iron, however, proved to be intractable to manufacture, and it was seven years before Hadfield's company delivered its first ton of transformer sheet. In the ensuing 17 years silicon iron saved the electrical industry about $340 million—an enormous amount of money in the 1920's.

The next leap forward in core technology had its roots in the early 1930's, when the American metallurgist Norman P. Goss of the Cold Metal Process Company found that combined rolling and heat treatment of silicon iron produced a sheet with outstanding magnetic properties in the direction of the rolling. Goss did not realize it, but the effect of the process was to align the major axes of the iron crystals in the same direction, producing a cooperative magnetic interaction. When a core made from such a material was oriented properly in a transformer, the saturation improved 50 percent, the hysteresis losses dropped by a factor of four and the permeability increased fivefold.

Again, the translation of that discovery into a method of production of satisfactory iron sheet was long and painful. The Westinghouse Electric Corporation and the American Rolling Mill Company teamed up to develop the processes, as did the General Electric Company and the Allegheny Ludlum Steel Company. Cross-licensing between the two groups enabled the transformer manufacturers to exploit one another's advances.

The requirement for a specific orientation of the metal in the core also necessitated substantial changes in the manufacturing of the core. No longer could a simple E form be stamped out of an iron sheet; in order to achieve optimal results, each leg of the E had to be made from a separate punching. Altogether, Goss's discovery did not become commercial reality until 1941, but its subsequent effect on transformer improvement was substantial.



and then cut the voltage even further for commercial or small-scale industrial uses and for residential consumers. Although this schematic depiction is linear, in practice power lines branch at each voltage reduction to establish the distribution network.

Also bearing on the transformer's performance are electrical insulation and cooling systems. These two systems are intimately related because the amount of heat the core and conductors generate determines the longevity of the insulation, and the insulation itself—whether solid, liquid or gas—serves to carry off some of the heat. Temperatures inside a transformer unit typically reach 100 degrees Celsius, the boiling point of water. Under such conditions deterioration of the insulating materials can limit the lifetime of a transformer.

Air provided the only insulation and cooling in the first Stanley transformers; the cotton that covered the wires served mostly to hold them apart. Soon afterward George Westinghouse immersed the entire transformer in a tank of oil and spaced the laminations in the core so that the oil could circulate by convection among them. The insulating properties of oil-soaked cotton turned out to be superior to those of dry cotton in air, and the combination of circulating oil and a variety of oil-impregnated cellulose materials such as kraft paper became a standard that is still widely employed today.

Although oils are inexpensive and effective as insulators and coolants, their flammability makes them unacceptable for units placed inside buildings. Chlorinated hydrocarbon liquids (PCB's) introduced in 1932 are not flammable and were once used extensively, but the recent discovery that such compounds have long-term toxic effects has prompted a ban on their use. Some transformers rely on air or nitrogen and glass-based insulators; these are essentially fireproof and can be installed indoors. The breakdown strength of the gas is sometimes enhanced by the addition of small quantities of fluorocarbons. Other dry transformers depend on cast-resin insulation made of polymerizing liquids that harden into high-integrity solids.

Technical progress in heat removal is largely responsible for reducing the overall size of the transformer assembly. Transformers insulated with oil relied at first on natural convection to circulate the coolant, but now rather elaborate means of removing heat from the oil have been devised. Many units have fan-cooled, external radiators through which the oil circulates by convection or pumping.

Engineers have also been experimenting for many years with vapor cooling: a nonconducting liquid with a low boiling point vaporizes when it comes in contact with hot parts, is transported as a gas to a separate compartment and condenses there. Several transformers that have vapor-cooling systems are in operation, but their cost is not yet competitive with that of conventional units. The technology still holds promise and is being actively pursued.

As the transformer enters its second century of service it is not easy to predict how its evolution will proceed. Research on amorphous metals (metals that essentially have no crystal structure) has elicited some very promising magnetic properties, but economical methods of producing such materials have yet to be demonstrated. Superconducting transformers, whose coils have no electrical resistance, have been built for demonstration in laboratories, but their temperature must be maintained at a few degrees above absolute zero and so they too are still impractical. Even though such experiments promise technical advances, the present state of the market, which is depressed because of overcapacity in the electric power industry, will probably delay any moves to change radically the way transformers are made, although gradual improvements in insulation and cooling will undoubtedly continue.

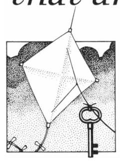
One might ask, finally: Is the transformer here to stay? Solid-state circuitry has greatly reduced or eliminated the need for transformers in small electronic apparatus such as radios, sound systems and television sets. And the availability of much larger solid-state devices has made it feasible in some cases to transmit high-voltage electric power as direct current rather than alternating current, although transformers are still required in the conversion process. These are hints, however, that solid-state devices could take over some of the jobs of transformers in power systems.

The recent breakthroughs in high-temperature superconductivity have raised hopes that materials might be found that are superconducting at room temperature. If they are, and if they can carry very large currents, the distribution of electricity as low-voltage direct current rather than as alternating current might become practical. Yet no one has any idea when or if such materials may be formulated, and it is far from clear that such a drastic change in established power systems would make economic sense.

In view of such advances and the unpredictable history of technological change, it would be foolhardy to maintain that the transformer will be here forever. But it seems very likely that the transformer will serve in its second century as it did in its first—silently, efficiently and unobtrusively supporting the electric power systems on which so much of modern life depends.

THE AMATEUR SCIENTIST

What explains subjective-contour illusions, those bright spots that are not really there?



by Jearl Walker

If you look at the illustration at the left below, scanning the entire pattern without fixing your gaze on a side of the broken triangle, you are likely to see an illusory opaque, white triangle that seems to hide parts of four black circles and an outlined triangle. The illusory triangle, which may take several seconds to appear, probably seems to be appreciably brighter than the white background. (You can sharpen the illusion by slipping a dark sheet of paper under the page to eliminate any interference from what is printed on the other side.)

The opaque triangle is an example of a class of illusions called subjective contours, which were first brought to popular attention in 1976 by Gaetano Kanizsa of the University of Trieste [see "Subjective Contours," by Gaetano Kanizsa; SCIENTIFIC AMERICAN, April, 1976]. Variations of the patterns that produce illusory triangles are now known as Kanizsa triangles.

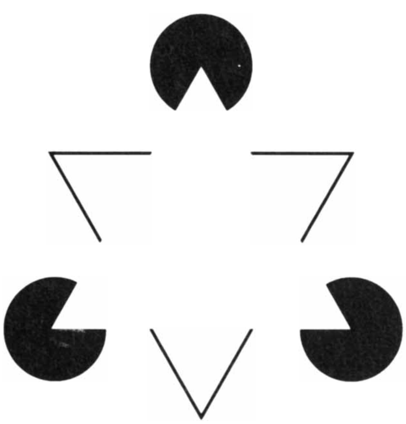
Why does the subjective triangle appear, and why is it brighter than the region surrounding it? These questions have resisted definitive answers for more than two decades in spite of the concentrated efforts of physical psychologists and workers in other disciplines. (Patterns that generate subjective contours are easy to construct, and you may want to try your hand at discovering new patterns or testing explanations of their origin.)

One early explanation involved the brightness contrast in the Kanizsa pattern. It maintained that because the corners of the white triangle contrast sharply with the incomplete black circles, the corners are perceived to be brighter than they really are and the brightness somehow spreads to affect the entire illusory triangle. That explanation was easily toppled. If the illusory triangle is outlined with a thin border so that the triangle is no longer subjective, the brightness disappears or at least is so weakened that it may not be noticed by a casual observer. Yet the brightness contrast is undiminished at the triangle's corners. The brightness-contrast theory also fails because it does not explain why the rest of the white regions bordering the incomplete circles are not also brightened, even though they contrast just as sharply with the circles as the "wedges" cut out of the circles do.

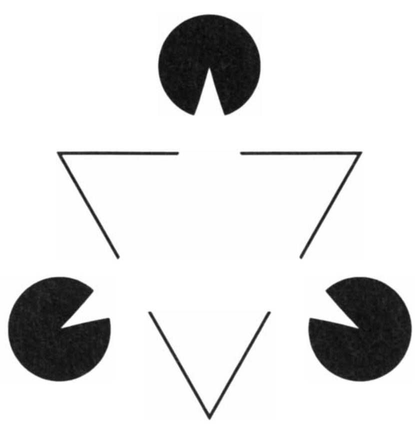
Another early explanation dealt with the parts of the visual system that are responsible for detecting lines, edges and orientations. Could they interact in some way to bring about the illusion of a bright, completed triangle in Kanizsa's pattern? As tempting as the explanation is, it fails to explain why subjective contours are sometimes more readily seen when the patterns are blurry or badly illuminated.

A different part of the visual system is assigned the task of detecting the repetition of such elements as lines, apparently through a process analogous to Fourier analysis. Might the analysis fill in enough extra lines to complete a subjective figure? Apparently the analysis cannot be the source of the illusion, because subjective contours have been produced in patterns that lack the repetition required to provoke such analysis.

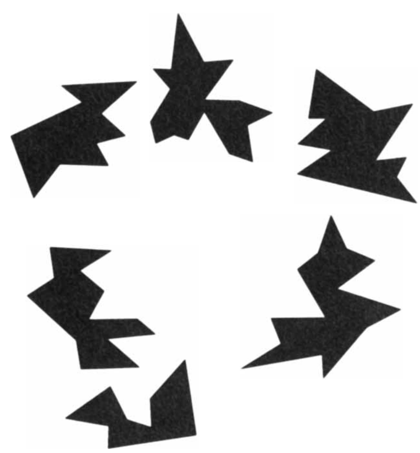
In 1986 Stanley Coren of the University of British Columbia, Clare Porac of the University of Victoria and Leonard H. Theodor of York University reviewed and rejected such explanations, which implicate one or another physiological source. They supported instead a second class of explanations, involving cognitive processes. You might, for example, perceive a subjective triangle because your brain automatically searches for ways to fence in regions and thereby make sense of an initial-



A Kanizsa triangle



A subjective triangle with curved sides

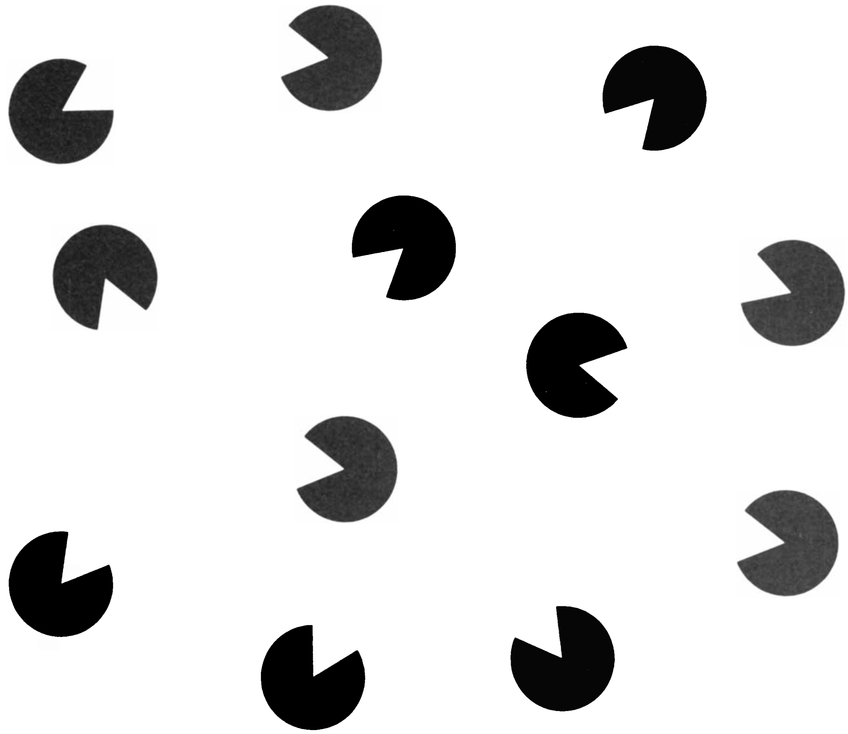


A triangle in a "noisy" pattern

ly strange pattern. Alternatively, the triangle might appear because you regard the pattern as a puzzle, which you must solve by perceiving familiar figures in it.

Coren had previously argued that one strong cognitive mechanism at work in patterns producing subjective contours involves their apparent depth. Notice that in the case of the Kanizsa triangle the subjective figure seems to be in front of the other figures, partially blocking your view of them. Why does the triangle brighten when depth is perceived in the pattern? No one can agree on the answer, but it may well be that the brightness helps you to rationalize your perception of the triangle, which you know is illusory. Coren had also put forward another cognitive mechanism that may have a major role in subjective contours. When you perceive an illusory figure in a pattern, you draw on earlier experiences even if you do not recognize that you are doing so. The body of previous experiences is called a perceptual set. For example, once you have learned to see the triangle in Kanizsa's original pattern, you then readily see a triangular figure with curved sides in the middle illustration on the opposite page.

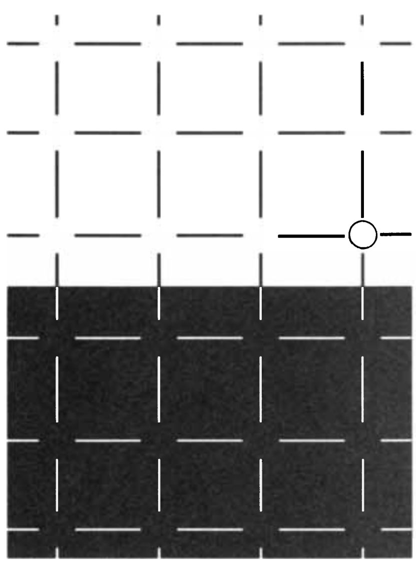
A sharper test comes with the third illustration on the opposite page, a pattern taken from experiments reported in 1979 by Irvin Rock of Rutgers University and Richard Anson, who was then also at Rutgers. The pattern is an extension of earlier work by Coren. After covering up the first two figures, ask someone unfamiliar with subjective contours to examine the third figure. The observer is unlikely to pick out an illusory triangle at the center of the pattern or to see any unusual brightening there. Then familiarize the observer with the first two figures. The chances are that a bright illusory triangle in the third pattern will now be recognized. The illusion may not appear for tens of seconds, however. If its source were physiological instead of cognitive, there would presumably be no delay. Can you find an illusory figure in the top illustration on this page, which is also adapted from work by Rock and Anson?



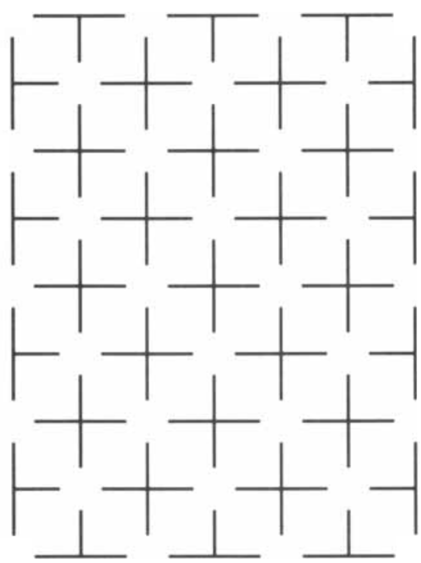
Is there any subjective figure here?

A different kind of illusion can be seen in the incomplete grids in the illustration at the bottom left on this page. The illusion was first investigated by W. Ehrenstein and is now named after him. If you avoid fixation, you will see a circle, square or blob at the points where the lines would intersect if the grid were completed. The figures are particularly bright in the grid of black lines and particularly dark in the grid of white lines. In both cases the figures probably appear to lie above the grid, seemingly hiding intersections of the lines. The illusion is just as apparent if you examine a single cell of the grid, with only one incomplete intersection. If you fix your gaze precisely on the region of intersection, however, the brightened or darkened spot there disappears. It also disappears if you add a circle around the intersection region. A striking variation of the Ehrenstein illusion can be seen in the illustration at the right below: in addition to bright spots at the regions of intersection you also perceive bright diagonal streets connecting the spots.

What accounts for these illusory



The Ehrenstein grids



The Ehrenstein "street" illusion

spots and streets? The spots might be produced by a desire to close parts of the pattern by mentally superposing something that connects the ends of the lines at each incomplete intersection. Alternatively, the spots might come about because of the contrast in brightness between the ends of the lines and the space at the incomplete intersection. A third possibility is that you mentally add depth to the pattern, transforming a flat drawing into a grid whose intersections are hidden by small spheres or—in the case of the streets—by another grid laid on top of the actual one. This explanation seems to me to be important, because when I examine the illustrations, I have a strong perception of depth.

In the late 1970's John M. Kennedy of the University of Toronto invented a variety of illusions that appear to be related to the Ehrenstein illusion. In the first of the four patterns in the illustration at the upper left below, 'a simple layout of radial lines generates the illusion that a bright circle fills the center of the pattern. Kennedy argued that the illusory figure is due to the contrast in brightness between the inner end of the lines and the white center. According to theory, the region just beyond the end of a line seems to be brighter than it really is because it is close to the dark line. In the case of a single line, the brightening is imperceptible, but if several lines are grouped so that the brightened regions are adjacent or overlap one another, the brightening is quite perceptible.

In the next three patterns in the group the lines are progressively reoriented, until they are finally tangential to the central region. As the reorientation continues, the illusion of a bright center weakens and then disappears. According to Kennedy, the reorientation of the lines shifts the regions involved in brightness contrast out of the central region. When the lines are tangential, the white space just beyond the end of each line is near another line; the space is not mentally brightened and the central region no longer exhibits the original illusion.

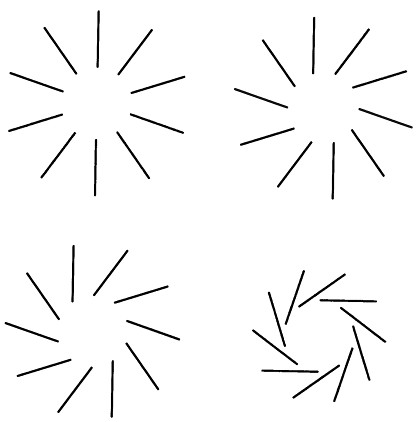
Critics have argued that the illusion is due not to a contrast in brightness but to the mental connecting of the lines across the central region and the perception of depth in the pattern: you imagine that the lines intersect but that the intersection is hidden from view by a bright sphere lying above the intersection. When the lines are reoriented, the idea of a hidden connection becomes less viable, destroying the illusion.

I can add an observation in support of this theory. If the illusion is due to a contrast in brightness at the end of each line, why then is the region just beyond the outer end of the lines not brightened? Even if I include many more lines in the pattern, so that the affected regions near the outer ends are closer together, I find no brightening and perceive no illusory figure outside the lines. The idea that depth and a hidden intersection must be perceived in the pattern in order to create the illusion of a central figure appeals to me. Everyone is accustomed to scenes in which an object blocks the view of more distant objects, and I suspect that I automatically bring such a perceptual set to bear on Kennedy's patterns.

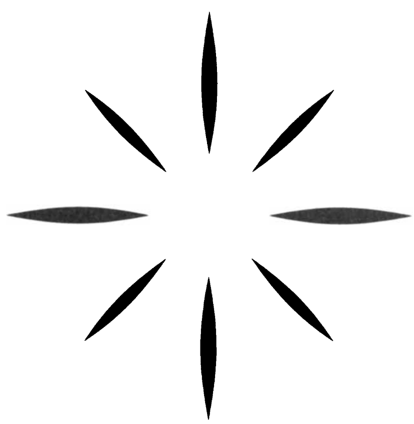
To counter the theory that a hidden connection is necessary to the illusion, Kennedy showed that the illusion remains when the lines are replaced with petal-like figures that are widest at the middle and pointed at the ends, as in the illustration at the upper right on this page. The pointed ends should remove the temptation to imagine a hidden connection. For me, however, the illusion of a central figure is considerably weaker with this pattern than it is in the case of lines, suggesting that although a hidden connection may not be necessary, it certainly helps.

Another pattern investigated by Kennedy consists of radial "lines" that are strips of black dots; the dots fade toward the inner end of each line. A version of Kennedy's pattern, devised by Barry L. Richardson of the University of Toronto, is seen in the illustration at the lower left on this page. It has solid but tapered lines. Both patterns create the illusion of a bright center resembling the sun. A photographic negative of either pattern generates a dark center, as if a black hole were there.

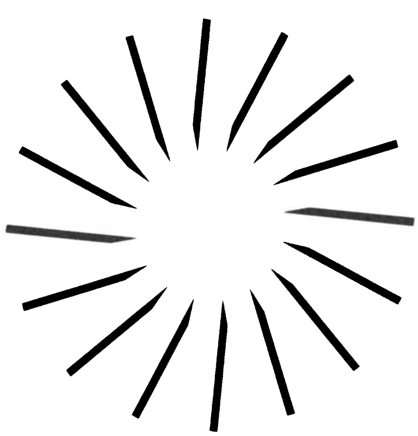
An illusory bright center is also produced by a Kennedy pattern that



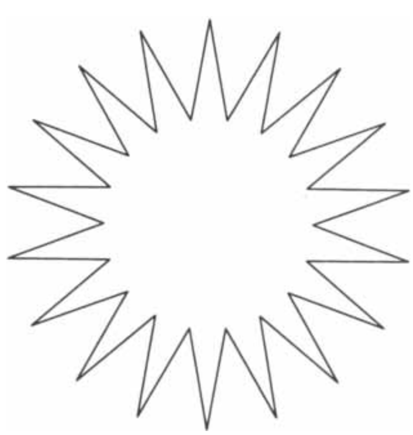
The effect of varying line orientation



The effect of petal shapes



The effect of tapered lines



A sunlike pattern

resembles a stylized sun [*see illustration at lower right on opposite page*]. According to Kennedy, the brightening is due to the contrast in brightness between the central region and the sharp bends formed by the line segments. This pattern is important because the illusion does not serve to connect or enclose anything in the pattern: all the line segments are already connected and the center is already fully enclosed. Are Kennedy's illusions due to brightness contrast, perceptual set, the illusion of depth and hidden connection or just a playful desire to see something interesting in an otherwise dull illustration? I leave the matter to you.

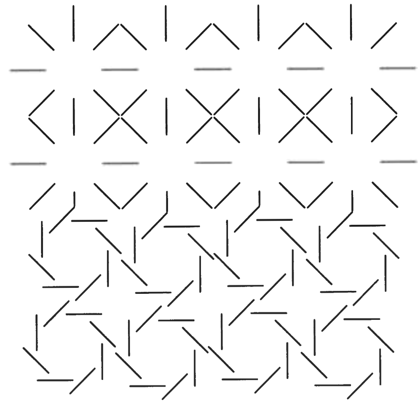
In 1983 Alex Stewart Fraser of the University of Cincinnati offered up an intriguing combination of Kennedy and Ehrenstein patterns. When you examine the top illustration on this page from a normal viewing distance, the top half produces bright spots as in the Ehrenstein illusion, whereas the bottom half fails to produce any illusion. Note that the bottom half is filled with lines that could be described as being tangential to circular spots like those in Kennedy's patterns. The spots are not apparent, however, and there is no unusual brightness. Now move away from the illustration. When you are at an appropriate distance, the bright spots in the top half almost disappear, and now the bottom half seems to be filled with spots that are somewhat darker than they should be. Squinting greatly improves the illusion.

The illustration at the right, also from Fraser, contains an embedded pattern. When the illustration is viewed from a normal distance, the pattern is almost impossible to perceive. If you move away from the illustration and squint, a slightly darkened figure appears. It is generated by the spots that are surrounded by tangential lines.

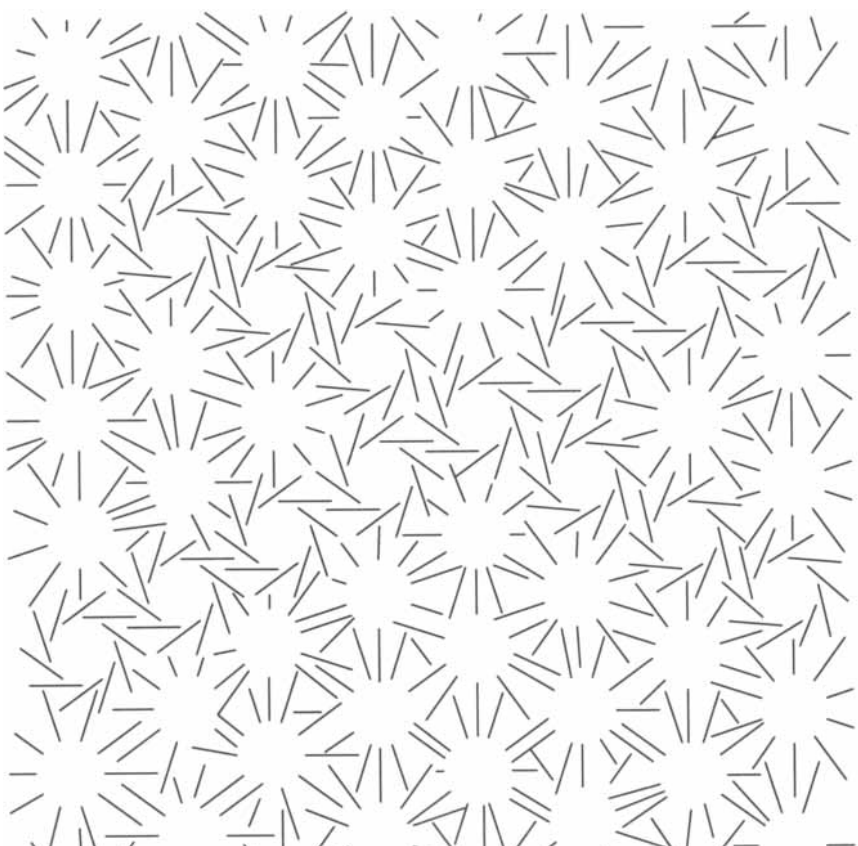
In 1981 Lothar Spillmann and Christoph Redies of the Neurological Clinic of the University of Freiburg, Germany, described how the basic Ehrenstein illusion can be modified if the grid is overlaid with a transparency of randomly placed dots. In the regions where bright spots normally lie, the dots seem to be less densely packed than they are elsewhere. They may even appear to be organized into concentric circles. To me these regions seem to bulge upward, as though they were hemispherical caps lying on the grid. The brightening that is normally produced by the grid is missing.

When the transparency is moved across the grid, the patches "boil" with activity and shift to new positions in the direction of the pull. The shift may be due to a persistence of vision associated with the patches. When the transparency is similarly drawn across an Ehrenstein grid that normally produces an illusion of bright streets, the streets are also shifted in the direction of the pull. If the transparency is pulled diagonally along one set of the streets, those streets come alive, whereas the perpendicular streets may disappear. Transparencies bearing either regularly placed dots or hatched or wavy lines fail to eliminate the normal brightening in the grid or, when they are slid across the grid, to shift the illusory regions.

A similar illusion can be produced if a transparency of the Ehrenstein street grid is placed on the screen of a television set tuned to an empty channel (a black-and-white set is best). Inside the illusory regions of the grid the random "snow" on the screen seems to swirl either clockwise or counterclockwise. The motion is related to other experiments with kinetic random-dot arrays that were described in this department in April, 1980.



A pattern devised by Alex Stewart Fraser



A Fraser pattern with an embedded figure

COMPUTER RECREATIONS

Nanotechnology: wherein molecular computers control tiny circulatory submarines

by A. K. Dewdney

Will machines the size of a mitochondrion someday be injected into the human body? If the thought of tiny machines creeping within one's flesh makes one's flesh creep, consider the possible advantages. Microscopic machines might be made to act as tiny vascular submarines that destroy unwanted organisms and fat in the human circulatory system. With such machines a person might be destined for a longer life span.

The minuscule healing machines are in part the brainchild of K. Eric Drexler, a visiting scholar at the department of computer science of Stanford University. For more than a decade Drexler has been laying the blueprints for nanotechnology, a revolutionary concept of machinery in the nanometer (billionths of a meter) size range. He has designed gears and bearings of atomic dimensions and described molecular manipulators. He has even laid out the logic for a nanocomputer, the centerpiece of this month's excursion. Although nanotechnology is currently hardly more than a gleam in the eye of Drexler and a handful of other scientists, it has a curious ring of inevitability. The future it implies seems an order of magnitude more wonderful than anything in science fiction.

Drexler is not the first to explore the territory of nanotechnology. In 1959 Richard P. Feynman envisioned a succession of machines built to progressively smaller dimensions; machines at one scale of magnitude would construct those at the next level of size downward. According to Drexler, the ultimate products would have moving parts no larger than a few atoms across. Such products, depending on design and purpose, might roam through the human body, invading cancerous cells and rearranging their DNA. Other machines might swarm as a barely visible metallic film over an outdoor construction site. In the course of a few days an elegant building would take shape.

Drexler believes a key step in developing nanotechnology is the creation of assemblers, small machines that would guide chemical bonding operations by manipulating reactive molecules. Programmed to carry out certain construction tasks, assemblers would turn out specific machines by means of miniature assembly lines. Every hour entire factories no larger than a grain of sand might generate billions of machines that would look like a mass of dust streaming steadily from the factory doors—or like a cloudy solution suspended in water.

There are already aqueous nanomachines (of a kind) in nature. Consider the well-known *T4* bacteriophage, which replicates itself by invading the body of a bacterial host. The phage attaches itself to the outer cell wall of the bacterium and then, like a tiny hypodermic syringe, injects its own DNA into the host. The phage DNA redirects protein synthesis inside the bacterium so that hundreds of minuscule phage parts are manufactured. Through random encounters the parts proceed to assemble themselves into new DNA-loaded syringes that, once the bacterial wall disintegrates, are free to search for new victims.

Another example of nanomachinery is already a part of some bacteria possessing a flagellum, a long, helical structure protruding from one or both ends of the bacterium. The organisms swim by rotating the flagellum, which whips about in a circular, corkscrew fashion. A secret motor just inside the bacterium's cell wall drives the flagellum. The motor consists essentially of a protein rotor powered by ionic forces.

A key device in any conceivable nanotechnology is the bearing. Bearings normally require lubrication of some kind. In 1959 Feynman noted that "lubrication involves some interesting points. The effective viscosity of oil would be higher and higher as we went down.... But actually we may not have to lubricate at all! We have a lot of extra force. Let the bearings run dry; they won't run hot because the heat escapes away from such a small device very, very rapidly."

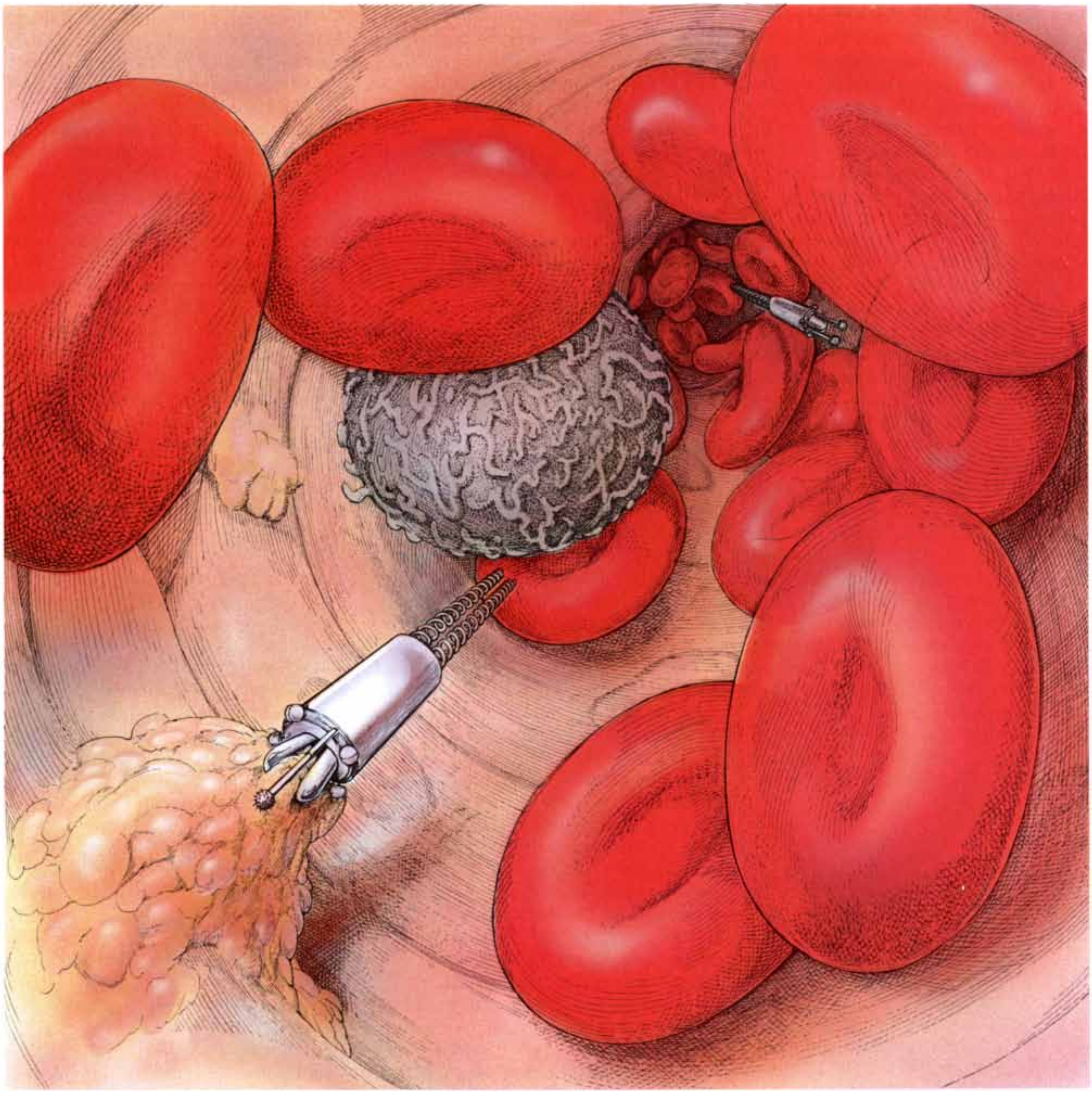
Would a lighter lubricant work? Even if kerosene were substituted for oil, viscosity would increase. As Drexler has pointed out, however, "from the perspective of a typical nanomachine, a kerosene molecule is an object, not a lubricant." In a paper devoted to gears and bearings, he addresses the problem of friction in roller bearings. Owing to what could be called atomic bumpiness, one might expect nanometer-size roller bearings to experience considerable static and sliding friction. Such bumpiness is described in part by a formula known as the van der Waals interaction potential. The formula contains a repulsive, exponential term that generates most of the bumps.

Drexler suggests that the friction problem can be handled by employing roller bearings in which surface atoms arranged in rows are meshed with similar rows in the race, the track in which a bearing moves. An assembler that can arrange single atoms of elements such as carbon and fluorine might construct the bearing and race from carbon atoms arranged in the single-bond structure characteristic of diamond. The bearing would thus be a sheet bent back on itself to form a hollow cylinder. The surface of both bearing and race would be studded with fluorine atoms arranged in angled rows [*see illustration on page 102*]. The same strategy could be followed to construct gears that mesh.

Drexler also addresses the problem of sliding friction between atomic surfaces. Given the ability to arrange atoms in positions that do not violate physical laws, one can exploit patterns that smooth out the van der Waals repulsive forces. For example, an atom that moves close to a surface consisting of rows of other atoms suitably staggered would experience a relatively smooth ride. But it must proceed along a straight path just above the groove between two rows. Drexler bubbles over with additional plans. Some are available only in technical papers, which Drexler has kindly agreed to make available to readers who want to know more about the speculative nuts and bolts of nanotechnology. The papers can be ordered from the Foresight Institute, P.O. Box 61058, Palo Alto, Calif. 94306. Readers who send $3.50 will receive a packet of four papers on various aspects of nanotechnology. A more general overview of the field will be found in Drexler's book, *Engines of Creation* [see "Bibliography," page 112].

Suppose for the moment that something like the miniature vascular submarine I mentioned above is actually possible. How will it be controlled? By a nanocomputer, of course, but of what kind? Electronic computers may be possible, Drexler says, but he has concentrated on mechanical computers based on what he calls "rod logic." Although their logical operations are mediated entirely by molecular rods that slide through a miniature matrix, they are not at all slow. At the atomic scale such rods take only about 50 picoseconds (trillionths of a second) to slide.

The principal element of a computer is the logic gate. One or more signal paths enter a logic gate and one path (which may later split off into other paths) leaves it. Signals are of two types, denoted by 0 and 1. Because the signal in a given path varies with time between two values, a



A nanomachine swimming through a capillary attacks a fat deposit

variable such as *x* can be used to label the path. A formula describes the logic function of the gate. For example, the formula "*x* and *y*" describes the so-called and gate; the gate's output signal is a 1 if and only if the input signals, *x* and *y*, are both 1. Similarly, the formula "*x* or *y*" describes the or gate. In this case the gate's output is a 1 if either of the input signals, *x* or *y*, is a 1. From these two types of gate, along with a third called a not gate, or inverter, one can build any logic function whatever. The not gate changes a 1 into a 0 and a 0 into a 1.

In current computers, logic gates are made out of micron-scale transistors implanted in the surface of a chip. A layer of specially treated silicon infused with impurities that carry an excess or a shortage of electrons is covered with a layer of polycrystalline silicon. A third, metallic layer, usually aluminum, overlies the first two. It acts as a conductor between transistors. Of course, not all points on the surface of a chip contain all three layers. The precise pattern in which the layers are deposited actually defines the transistors and their interconnections. The metal layer, for example, consists of ultranarrow aluminum strips running helter-skelter all over the surface.

In a microchip, signals that travel the tiny aluminum paths vary between two distinct levels of voltage that encode the logic values 0 and 1. The gates in general are nand gates, which, in terms of the simple logic functions described above, can be represented by the expression "not (*x* and *y*)." In other words, the nand gate has an output of 1 when it is not the case that both inputs *x* and *y* are 1. Thus if either input is 0, the ouput is 1. One can build any logic function, no matter how complicated, from nand gates alone.

Microchips are small. Their components are scaled in the micron (millionth of a meter) range. Drexler now asks us to consider a computer that would fit inside a single silicon transistor!

The logic in Drexler's hypothetical nanocomputer is mediated by infinitely small rods instead of wires holding particular voltage levels that encode 0 and 1. A rod of atomic dimensions is free to slide into one of two positions. The positions of the rod encode the values 0 and 1 by convention. The rod is made of carbyne, a chain of carbon atoms that are linked by alternating single and triple bonds. Carbyne turns out to be strong and stiff enough for the job, yet it can be bent around gentle corners and still slide. Along their lengths the carbyne rods sport knobs. The knobs come in two varieties [*see top illustration on opposite page*]. Some of the knobs simply project outward, and they are called probe knobs. Others look like panels or doors, and they are called blocking knobs.

The base of both kinds of knob is formed by a hexagonal pyridazine ring, linked on either side to the carbyne chain. A probe knob consists of a second carbon ring above the pyridazine ring. The superior ring holds a single fluorine atom in a forward position, ready to act as a logic probe. A blocking knob differs slightly from a probe knob in that it is surmounted by a fluorinated thiophene ring. But names are not important. We assume in what follows that such structures can handle the jobs of probing and blocking. Suffice it to say that Drexler has calculated much more than I have room to discuss here, particularly the amount of tension the rods must have to endure thermal jostling.

Besides its rods, the logic heart of the mechanical nanocomputer consists of a three-dimensional matrix made of atoms near carbon in the periodic table. It is a framework penetrated by channels that run in two orthogonal directions on a great many levels. How do the rods compute logic functions within the matrix?

Imagine a single, horizontal rod resting in its channel in one of the two allowed positions. Called a logic rod, it will, for the sake of simplicity, have two probe knobs and one blocking knob [*see bottom illustration on opposite page*]. The rod therefore passes through three locks, hollow parts of the matrix in which two rods (and their knobs) meet at right angles. For a given lock the question is whether the probe knob will slide easily past the blocking knob or be held by it. Of the three locks the logic rod in our example passes through, the two leftmost ones contain blocking knobs that are attached to vertical rods labeled *x* and *y*. The lock at the right contains a probe knob, which is attached to a third vertical rod labeled *f*.

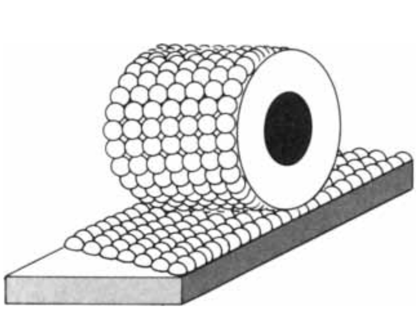
The horizontal rod and its locks serve to compute the nand operation. The two vertical rods at the left encode input variables *x* and *y*. The vertical rod at the right represents the output of the nand operation.

If one supposes that a vertical input rod encodes 0 in the up position and 1 in the down position, it may happen that *x* is 0 and *y* is 1. In other words, rod *x* is up and rod *y* is down. In such a case it is impossible to pull the logic rod leftward past both vertical rods because rod *x* is in the up position; its blocking knob gets in the way of the logic rod's probe knob. Only if rod *x* is also down will both probe knobs of the logic rod clear their respective blocking knobs so that the logic rod can slide to the left.

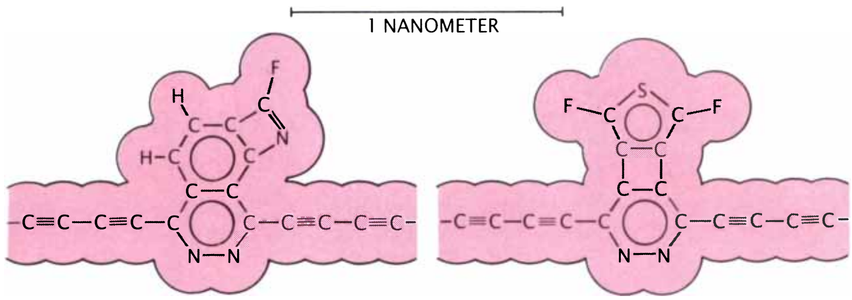
What about the output rod *f*? It may test whether the logic rod has found its new position by pushing into the lock containing the logic rod's blocking knob. It happens that when the horizontal rod is in its right-hand position, rod *f* will clear it, going from logic value 0 to 1. But if the logic rod is able to slip to the left, its blocking knob prevents rod *f*'s probe knob from moving through the lock, and so *f* has a value of 0. Thus the value of *f* must be 1 if *x* and *y* are not both 1, in keeping with the rules of the nand function.

The computational cycle of Drexler's rod computer involves three distinct stages that follow one another with lightning speed repeatedly. In the first stage the input locks are set by temporarily pulling on the ends of input rods. This allows all the logic rods to be withdrawn to their initial positions. In the second stage the input rods take up their current positions. They might be connected to various sensors of the nanomachine or to outputs of the nanocomputer itself. Now a source of displacement tugs all the logic rods to the left, so to speak. In the final stage the values of output variables are decided when the output rods are all pulled. The rods that can move do so.

Investigators working in artificial intelligence are sure to raise objections at this point; granted even that adequate computing power can be



A roller bearing of fluorinated diamond



A probe knob* (left) *and a gate knob* (right) *on a carbyne logic rod

built into such a small package, how does Drexler propose to incorporate the requisite intelligent behavior into his little autonomous vehicles?

Drexler replies that in spite of the fact that quite powerful central processing units might be embedded in quite minute spaces, some of his conceptual nanomachines do not have to be terrifically smart in any case. Take the nanosubmarine described earlier, one of the engines of healing. The submarine travels through the bloodstream, absorbing glucose and oxygen as it goes. Those chemicals supply it with power. Presumably someone hosting a fleet of billions might feel a bit peaked from time to time. In any event, each nanosub uses the energy from its glucose engines (which are currently being designed) to drive two helical propellers. It uses no guidance whatever. When it bumps into something, it is notified by the nano equivalent of a contact switch in its forward sensors. The same sensor then tries out a variety of molecular fits by matching templates against patterns built into the sensor.

If the submarine detects wall cells, red cells, leukocytes or a finite number of other benign entities, it reverses engines briefly and then continues on its way. But if it encounters a hostile bacterium or virus, it opens its mechanical jaws to engulf what it can of what it has found. This includes fat deposits and virtually anything else it is programmed for [*see illustration on page 101*]. Whatever is engulfed is broken down and released into the bloodstream for removal by the kidneys. The chances of unwanted bodies' surviving for more than a few hours in such a busy bloodstream would be small indeed.

Of course, it is all a dream—for now, at least. We are barely at the threshold of a possible plunge into the nanorealm. Can Drexler possibly be right in asserting that something close to the dream could actually take place? He is used to defending his ideas and on occasion altering them in the light of serious objections from fellow scientists. He remains, however, a technologist without a technology, an informed speculator whose greatest contribution might be to stimulate the dreams that guide our technological development. As such he is not among the gung ho. The dangers of abuse make atomic weapons and genetic runaways look tame.

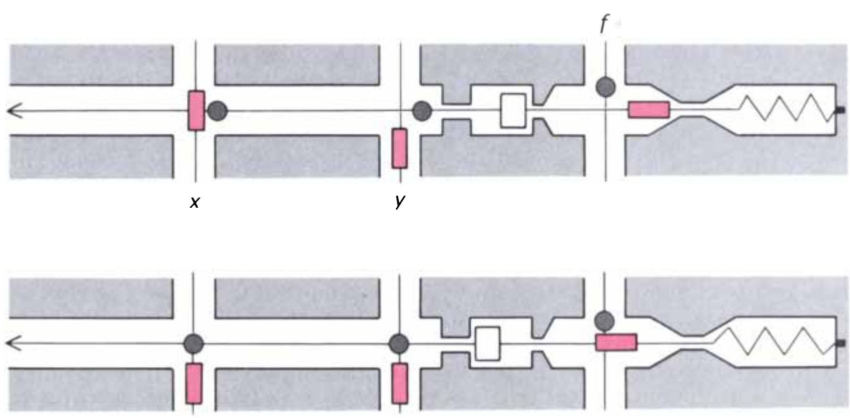
I was recently told by a nano enthusiast that "thousands now living will never die!" Strange to relate, an evangelist once made the very same pronouncement to my father half a century ago.

The September column about Rich Gold's party planning program involved a set of hypothetical guests invited to a party in a personal computer. The scene of the party is a grid of squares on the display screen. The guests move about from square to square in a search for happiness. Each guest has a preferred distance from each of the other guests. He or she will take a step in the direction that minimizes total unhappiness in the current configuration: the summed differences between current and preferred distances measure unhappiness.

The basic program cycle was the simplest one possible. Guests moved one at a time. According to Arthur Charlesworth of the University of Richmond, such simplicity can distort festivities somewhat. Charlesworth has been studying a parallel programming utility called the multiway rendezvous. If each guest were represented by a process in a parallel programming language, at the end of each iteration "all such processes would hold a 'meeting' to share the information about the new positions of guests." The distortion can be removed in our still commonplace sequential milieu by not moving the guests until the unhappiness of each has been computed.

Arthur D. Penser of Huntsville, Ala., was unhappy about the unhappiness function. He prefers parties in which the differences between current and ideal distances are squared. Given the same invitees in the same initial positions, how different would such parties be? Clearly, they would be more strongly pulled by extreme differences. Absence would make the heart grow fonder squared.

As I stated near the end of that column, there is a science of sorts called proxemics. William Ickinger, Jr., of the University of Maryland once constructed a hexagonal grid that could be laid out on a floor to study interpersonal distances at controlled gatherings of people. Ickinger is fascinated by the subject of programs that simulate gatherings, and he would like to correspond with anyone who shares his interest. He lives at 430 Ringwald Street, Chillicothe, Ohio 45601.



A logic rod defines the nand operation

BOOKS

The celestial suburbs, the trackless taiga, iron-bound islands, and humans—lean and fat

by Philip Morrison

EXPLORING THE SOUTHERN SKY: A PICTORIAL ATLAS FROM THE EUROPEAN SOUTHERN OBSERVATORY, by Svend Laustsen, Claus Madsen and Richard M. West. Springer-Verlag ($39). **NEARBY GALAXIES ATLAS,** by R. Brent Tully and J. Richard Fisher. Cambridge University Press (paperbound, $59.50). **THE SKY AT MANY WAVELENGTHS,** by Christine Jones and William Forman. Eleven color slides in folder, AS311, Astronomical Society of the Pacific, 1290 24th Avenue, San Francisco, Calif. 94122 ($16.50).

To celebrate "the extraordinary beauty of the southern sky" and to offer a sense of what it is they do, three Danish optical astronomers at the European Southern Observatory have spread a treasury of images over the big pages of this picture book. The domes and dorms and airstrip of the ESO station in the high Atacama Desert in Chile make up a solitary little mountain town that spreads over its treeless saddle at an altitude of 8,000 feet. The air is remarkably clear, stable and dry—two inches of rain a year—right for big telescopes and small, and for the large submillimeter-band radio dish just put in use.

Because the dazzling center of the Milky Way passes directly overhead, the black night blazes with clouds of stars. These latitudes offer the best view inward toward the center of our own city of stars, the Galaxy, and at the same time a fine view outward to the celestial suburbs we call the Magellanic Clouds. The ESO's outpost at La Silla serves close to half of the world's astronomers, particularly those from its eight member countries. Fourteen telescopes are now at work, the largest a fine 140-inch, a dozen years past its first light.

The big photographs are fresh and very well documented. This is a run of vintage seasons for southern skywatchers; we are shown what they saw of Comet Halley in 1986, particularly its changing gas tail, in quite fine views. Comet West (found by the third author in 1976), trailing its multiple dust tails, was far brighter and more colorful; it has fled back into space and will not be seen again for a million years. Thanks to the authors' energy their intricate atlas shares in the February 1987 sensation: Supernova 1987A. No supernova has flared in such close proximity for 300 years. The star floats in the bar of the Large Magellanic Cloud, shown here in before-and-after color shots, amidst an entire gallery of bright red emission nebulas and a superposed foam of bubbles, probably the remnants of other spent supernovas from the distant past, with which our visible guest star is somehow involved.

The Milky Way is, as it ought to be, the centerpiece of this atlas. A bonus inclusion consists of a yard-long panorama of the full circle of the Milky Way, assembled from eight photographs of the sky made with a small, guided wide-angle camera. Some of the pictures were shot from La Palma in the Canary Islands, some from the La Silla station. The authors offer a guided tour of the disk, commenting on many features of the Galaxy, which they amplify through deep and detailed photographs of the regions. Three big, exciting pages, for instance, show rich fields at the very center of the Galaxy. One field is viewed through a heavy overcast, much intervening dust, right in front of the distant hub, whose presence is revealed by radio and infrared emissions. In adjoining fields the clouds dwindle, until myriads of stars offer evidence that we are looking beyond the dust, to see the far-off edge of the great starry central bulge of the Galaxy. Here is that wonderful optical window found long ago by Walter Baade, a dust-free opening through which he could study stars that are nearly 10 times as close to the galactic center as the sun is.

External galaxies are also presented: there is a bright little smooth-textured elliptical galaxy, apparently now the center of a big flat disk of stars like that of a spiral galaxy, and a stunning view of the famous colliding pair in Corvus, the ones with long, starry antennas.

One small picture in pseudocolor is genuinely instructive. It shows a galaxy in two views made by a sensitive electronic imager, a CCD. In one view a red filter was used, and the image is printed in red. The other view used a blue filter. Superposed, the images show a galaxy in blue, decorated with dozens of fuzzy red spots, distributed mainly along the inner edges of its spiral arms, where brilliant young stars light up clouds of atomic hydrogen. Two nearby fields are expertly chosen to make a dramatic contrast. In one we look out to the empty depths of the sky, far from the Milky Way band. Only a few foreground stars are seen, along with a sparse array of distant galaxies; it is a grand void. A slight shift of direction fills the field of view with a nearby cluster of diverse galaxies in the constellation Fornax. The next page magnifies one of those small galaxy images into a detailed and colorful pinwheel.

The second atlas is pure cartography: it presents only large, clear maps that sum up the long optical and radio study of our extragalactic neighborhood. The first maps (all are the handwork of Jane J. Eckelman of Manoa Mapworks, Honolulu) plot with care the place in the sky of about 2,400 galaxies. They appear as circles, triangles, squares in black, blue, red and other colors, each galaxy coded for its apparent size, type and red shift. The band of the Milky Way is nearly devoid of external galaxies, of course, since its dust hides the background. The cap of the south galactic pole discloses a local void; there are relatively few nearby galaxies out that way. Northward toward the crowded Virgo Cluster the galaxies abound.

R. Brent Tully is an optical astronomer at Hawaii, J. Richard Fisher a radio astronomer at Green Bank. They have jointly pioneered powerful techniques for measuring galactic distances. The first set of maps, close to the data, are two-dimensional. They are simply projections onto the sky of clouds, clusters, spurs, voids. They do, however, make possible recognition of correlations in space: similar red shift can be taken as a mark of similarity in distance.

The authors name a couple of dozen evident groupings among these nearby galaxies, then undertake a bolder task. They pass from two dimensions and the red-shift hint to a full three-dimensional model. The maps in space begin with a map showing only colored dots on the sky's plane: the starting point. Now the red shift is taken seriously as being proportional to distance in space. The ambiguity introduced by internal motions within groups is recognized by demanding that groups remain usually rounded in space, without the bias of elongation in the line of sight. By plate 21 there is a composite of composites. We see a set of colorful smoothed contour maps, pools and coarse filaments of galaxies in space, plotting in greens, yellows and reds the density of galaxies on the sky. What comes out is the local supercluster: two thick layers, galaxies by the thousands, the larger layer to the north, the smaller to the south, of the plane of the Milky Way.

The projection used to set forth these dissected three-dimensional views from vantages far from home is less self-centered. It is based on the flattened local supercluster, recognized and described to a generally doubting community during the 1960's and 1970's by Antoinette and Gerard de Vaucouleurs of the University of Texas. Today that structure has been made the cornerstone; its principal plane holds the Virgo Cluster and the Local Group. The geometry of this inner neighborhood, nominally a quarter of a billion light-years across, is viewed this way and that in projections of a dozen bricklike domains. Two final maps extend tentatively and incompletely to superclusters at a tenfold distance. For readers who like to know just where they are and who make a genuine effort at visualization, this is a remarkably imaginative presentation. A separate catalogue volume presents a compilation of the galaxy data. To cite the authors: "Enough of words. The maps are more eloquent."

The third of these visual explorations is the work of two astronomers, Christine Jones and William Forman, who look more steadily with detectors in orbit than with eye at night. The authors have chosen the familiar 35-millimeter projection slide rather than the printed page as their medium. They have compiled from the contemporary research literature up-to-date and reliable charts. The maps present in comparable format first the entire sky in a careful painting as a keen and patient eye sees it, with the Milky Way as visible backbone; then the sky is rendered as it appears to all our newer senses. Those novel panoramas are shown as color-coded whole-sky maps made at wavelengths no eye can see: broadband radio, two distinct radio line emissions, two views in the far infrared, two more by X ray and one by gamma rays. (For that most energetic of well-explored radiation bands, the entire map was compiled from 200,000 photons, each one individually counted.) Last comes the deepest sky of all, the relict microwave glow of the big flash. An auxiliary slide shows which radiation from all spectral regions can pierce the atmosphere. This set opens wide today's windows on the universe (neutrino astronomy has yet to begin any map of the sky); of course, the tale is too rich to exhaust in the helpful few pages that accompany the slides.

The Natural History of the USSR, by Algirdas Knystautas. McGraw-Hill Book Company ($29.95).

From Finland to the Bering Sea the map of vegetation zones in the U.S.S.R. shows a green belt about 500 miles wide and more than 6,000 miles long; this is the taiga, the largest forest in the world, a featureless sea of green, dominated in most places by a few species of conifers, home to wildlife forms often known worldwide. Its lack of diversity is ascribed to its rather recent evolution. In the wetter European area the tall forest is dark and gloomy under evergreen spruce and fir, particularly toward the south. Across the Urals the skies are clearer; those Siberian lands receive only from six to 12 inches of rain a year. There the forest itself is lighter, more open; its trees are pines and particularly larches, undemanding trees adapted to growth on sandy or stony soils, supported by spreading shallow root systems able to seek out surface waters above the universal permafrost. The undergrowth is rich among the shed needles of the deciduous larches. The dry continental winters are bitterly cold, with little snow. On a hot summer day the temperature can climb into the high 90's, only to fall that night to a chilly 40 degrees.

"The taiga has a severe beauty all of its own, and is full of memorable sensations: the exhilarating smell of the pinewoods on a spring morning; the drumming of a Black Woodpecker carrying half a mile or more;... the tracks of a Brown Bear imprinted in the moss;...the mighty Siberian rivers as they roar and tumble through the forests...." One photograph from the air at low altitude offers us the taiga as a sea; the trackless green forest, its texture unvaried except for a few overgrown fire-caused clearings, stretches to the horizon. The broad blue Stony Tunguska River flows through the greenwood, a dirt road along one of its narrow banks, to merge its waters with the broader Yenisei, bound 1,000 miles north to the polar sea. The taiga is not without its events, whether topical, such as the recent visit of the congressmen to the radar station not far south of this scene, or cosmic. One summer morning in 1908 a comet flew out of space to burst prodigiously in the air only a couple of hundred miles upstream from the place of this empty view, felling a wide forest yet killing no one.

The taiga wildlife is also displayed for us. Its most typical birds are two closely similar species of large grouse called capercaillie. They are impressive birds: a male can weigh as much as a turkey. To come on one and start it unexpectedly into explosive, whirring flight is a "heart-stopping" experience. The group courtship displays of the western capercaillie are so absorbing to the birds that they even ignore shots fired close by. Many of the other animals most typical of the taiga are familiar to Americans and Canadians: the largest is the brown bear, grouped by some zoologists into a single species with our big grizzly and Alaskan brown. The largest of the deer kind is the moose; that same creature is to be seen browsing in wetlands and clearings for new shoots and high grass all the way westward from Maine and New Brunswick past Alaska to Scandinavia. The Siberian taiga remains the stronghold of the wolf pack, "magnificent predators," the author calls them. Everywhere there are red squirrels, wolverines, foxes, voles, hares, similar to or even the same as the forest forms of North America.

Although the taiga is so wide, the largest country in the world still has room for much variety. The Soviet lands have been classified into 19 distinct zones of life, from empty arctic wastes to hot subtropical desert. This young naturalist-author, a freelance wildlife writer and photographer, has for half a dozen years now made field trips with his team of photographic colleagues to many regions of the U.S.S.R., far from their home in Vilnius in the Lithuanian S.S.R. They have produced a variety of fine pictures and a lively narrative of lands and creatures. Perhaps anyone would expect to see polar bears swaying across the ice, or the splendid Siberian white tiger. (That biggest of cats is not at home in the taiga; it is an inhabitant of the richer mixed forests of Ussuriland, lying north of Vladivostok along the Sea of Japan.) Fewer readers will have been ready for the cobra or the big monitor lizard of central Asia, the giant fennel or a stately lotus.

Here also are the familiar public concerns and private pleasures of the wildlife world. Protective legislation and public determination to end the damage to habitat by industrial development are trends that have grown during the past two decades or so in the U.S.S.R. The Soviet and international Red Data Books of endangered species are cited frequently; nature reserves, some of them big new ones, are mapped and described in this book over just about all the zones. One of the oldest such Soviet nature reserves includes the remarkable delta the Volga River forms as it enters the landlocked Caspian Sea. There great flocks of breeding birds have found protection ever since the lands were set apart in 1919 through a decree issued by V. I. Lenin. On the side of pleasure, Algirdas Knystautas offers us with enthusiasm the very first color images of the rare wading ibisbill, photographs he himself made in 1986 high in the wild Tien Shan mountains.

The readers to whom this well-made volume appeals most intensely will immediately recognize their own tastes and make their way to a fine and unusual book.

ISLANDS, by H. W. Menard. Scientific American Library, distributed by W. H. Freeman and Company ($32.95).

First, find your island. Just how to do so is the topic of the fresh opening chapter of this final delightful work by the late author, distinguished marine geologist at Scripps whose handful of popular books are treasures for the general reader. Who else would begin with a clever piece of operations research that discloses pretty convincingly from the number of unexpected islands found per ship per decade that the islands of the Pacific were come on by the European explorers quite at random? The Polynesian discovery, itself probably made both by chance and by voyages of design, was "pigs, chickens, and all... the greatest maritime feat in human history."

In exploration it is not a small achievement to do "as well as pure chance." Striking photographs help us to grasp the circumstances at sea. One picture shows the orographic clouds that stand high every day as a white beacon over Palau; another shows the formidable cliffs of St. Helena, one of the islands called ironbound, that is, circled by towering cliffs beset by great breakers. That island stood high enough to have been spotted in 1502 at a distance by mariners tacking far eastward to seek favorable tradewinds.

The last chapter treats just as freshly the question of how life including human cultures reaches islands. The weapons of World War II were not able to end the long lives of the atolls, although combat littered the blue lagoons and white beaches with rusting bits of heavy equipment. Even thermonuclear tests make only small if lasting changes; an islet or two has gone into vapor, but the atolls "looked little the worse for wear."

Oceanic islands (not Long Island or even Ireland) are both interesting and important to geology and biology because they are in effect small, rather new, laboratorylike samples, near enough both in isolation and in uniformity. Young Darwin first glimpsed something of evolution in the strange limited fauna of the Galápagos, and he came to his famous inference about coral atolls from the charts alone, because they were simple enough to show the work of "a single geological factor—subsidence." In the same spirit, H. W. Menard uses the islands, high and low, to draw out and to apply the insights of plate tectonics, both in worldwide comparisons and statistics and in such regional patterns as the long line of the volcanic Hawaiian Islands and their submerged cousins. He builds in this inductive way the most understandable survey of that theory to be found in the popular literature. Eye-catching support from maps, graphs and photographs is a large part of his success. Here are the islands of the world: Iceland, Bouvet, Tahiti, the Canaries, Réunion and dozens more, pictured and discussed not merely as varied specimens of beauty and curiosity but as instances of a single grand process.

One dramatic page compares the silhouettes of tiny landlubber volcanoes such as Vesuvius with Mauna Loa and other great marine shield volcanoes. A spreading sea floor contracts and sinks slowly as it cools, so that across the Pacific volcanoes subside steadily. By a model calculation it becomes plain that mere random outpouring of volcanic material at the present observed rate could have built the islands we now see on the Pacific plate, scattering them sensibly according to the age of the crust they stand on.

It is the Hawaiian Islands that best expose the life course of an oceanic volcano. The first stage is youth, the fluid outpouring that built the great smooth shields of Hawaii. The second stage is old age; a stiffer lava pours out a modest "warty cover" onto the smooth shield. Then follows long inactivity and with it deep erosion; in a brief rebirth a small dramatic feature such as Diamond Head can appear to adorn the long-cold shield with a finial cliff or two.

The simplest regularity is the steady aging of the islands of the Hawaiian group in sequence as one passes along the line of islands toward the northwest, as the crustal plate is dragged over a virtually stationary hot plume that wells up from below. Modern potassium-argon dating yields a plot of age versus distance from Kilauea that makes the point at once. Yet the exercise confirms what is implicit in facts determined in 1927, when the geologist Chester K. Wentworth dated the islands from fieldwork on erosion. (Menard has recalculated the case with more modern maps, to find an even closer fit.) Begin with an island of volcanic rock modeled to the original smooth shield form and wear it away to the present volume as mapped. The valleys and ridges along steep green once conical slopes are famous now, bedecked by waterfalls. Hawaii is high, smooth, still active and half a million years above water; distant Kauai's knife-edge ridges and steep headlands are 10 times older. That spectacular relief, dear to the tourist brochures, is not a sign of any special processes of erosion but merely of steady rainfall and conical shape. Wentworth even found one place where he could calibrate his relative measure of erosion time island by island. On Lanai the trunks of the old kiawe trees on the shore were buried up to a meter deep. The volume of sediment removed upslope could be estimated, and it was recorded that the foreign forest species had been introduced in 1837. The result fits the current radiometric ages within a factor of two. But at the time no one noticed Wentworth's powerful argument for seafloor motion.

Wise with hindsight from magnetic stripes and midocean ridges, we can now almost watch the Pacific plate moving by, all the way to the last long-submerged seamount of the Emperor Guyots up by the Aleutians. We even see the big plate shift direction; the rising hot plume anchored deep in the still core of the convection cell far below the ocean floor left its prints on the crust in the form of the island chain. The book ends with a few wondering paragraphs: Do the plants and animals that have survived and changed for 100 million years on this line of shifting islands show any signs of their long migration up the volcanic steps of a down escalator?

HUMAN BODY COMPOSITION: GROWTH, AGING, NUTRITION, AND ACTIVITY, by Gilbert B. Forbes. Springer-Verlag ($66).

Two commonplaces underlie this painstaking technical monograph. First, all life is watery, although the water within living forms is almost never fresh but as salty as the sea. Second, oil and water do not mix. A third more subtle biochemical result is also called on: almost every cell actively preserves the constancy of what seems a minor chemical difference. The saltwater within is rich in potassium, more than tenfold as abundant there as the similar ion of sodium. But outside the cell membrane, whether in blood plasma, in lymph fluids or in the open sea, that ratio is inverted.

It follows that a measure of the total potassium content of a living animal, say a fellow human being, is a good measure of the mass within its cells, the lean body mass, "the working, energy-metabolizing portion of the human body." Blood, like all extracellular water, is low in potassium, and above all the fat stored copiously as fuel reserve in tiny globules within fat cells is almost free of water and hence of dissolved potassium. Thus total body potassium measures active "body mass minus ether-extractable fat."

The amount of almost any element in the body can be measured by a variety of more or less invasive techniques. The most general method consists of injecting a known quantity of an isotopic tracer, allowing enough time for mixing and sampling for that element to check the dilution. Alone among the constituent atoms of life, the body's potassium can be assessed externally. Potassium has a rare natural isotope with a weak but detectable radioactivity, penetrating gamma rays that easily emerge from the body, intrinsic to life from its earliest origins. In order to measure body potassium, it is enough for the subject to bathe, don cotton pajamas and sit comfortably for a while in a chair near a large scintillation crystal, the business end of a gamma-ray counter.

The chief inconvenience is a little claustrophobia, for the sensitive instrument must be heavily shielded. (A television monitor playing in the room is often a practical palliative.) Gilbert B. Forbes's own counting vault in Rochester, its eight-inch steel walls lined with lead sheet, weighs 42 tons. There are 50 to 100 such whole-body counters of various designs—some are infant-sized—in use the world around. They were mainly built after 1960 for diagnostics and research into the problems of radioactive contamination. A normal adult will rack up 100 potassium counts per second in a good counter; in the days of atmospheric bomb tests and their widespread radio-cesium fallout the experimenter had need for good discrimination against that background.

The reward of such a gentle technique is the physiologist's ability to study not only a test series of a couple of dozen subjects but also a wide sample of humanity, by now tens of thousands of people, in sickness and in health. The full diversity of the human body is tested with respect to one simple quantitative partition: how we differ, by fat weight and by lean. On the average, young American men weigh in with a body-fat content of about 30 pounds, some 18 percent of total weight; young women average about 28 percent. The fat fraction grows steadily with age. There are data here on the fetus, infants and the very old, on hefty Sumo wrestlers and on lithe gymnasts, on seven-foot basketball players, on the anorexic and on the grossly obese. It looks as though lean body mass has an upper limit: an obese man who weighed more than 650 pounds had a lean body mass of about 180 pounds. The curve of lean mass against height for the towering basketball players flattens strongly toward a limit. Tentatively that limit is under 220 or 230 pounds for males and under 150 pounds for females; probably only by the use of steroids can the limit be broken.

The thin person proverbially held within every fatter one is here placed on the scales. One CAT scan shows an X-ray cross section of the mid-arms of two people who were matched in lean body mass although they were much different in total weight. The two dense islands of muscle with their central bones look quite the same; around the muscle and bone the thin arm shows only the thin border of skin, whereas the fatter arm is wrapped in a wide lagoon of low-density tissue.

The author is an expert participant in this line of study, the body simplified as two compartments. A concluding page offers one schematic summary. The center of the diagram represents the weight of the subject (generally an adult human); two distinct weight differences, in lean weight and in total weight, are plotted along the horizontal and vertical axes. Ten such lines radiate out in as many directions, their length roughly indicating the magnitude of change. One long vertical drop is quite exceptional: pure fat loss, with no change in lean body mass. It represents the hibernating bear, which has solved the problem uniquely. It loses considerable body fat but no lean mass over 60 days of winter sleep, slowly exhaling the products of fat combustion but able to reabsorb water and its solutes through the bladder walls. Migrating birds, fierce burners of fuel in their exhausting journeys, lose like exercising human beings both lean and fat; they follow the rule that the fatter the subject, the less the sacrifice of lean body mass in overall weight loss. Matched for sex, stature and age, obese people show a higher lean body mass than normal subjects.

A line toward two o'clock records overfed normal subjects; they gain mostly fat but some lean mass too. Normal pregnancy and the repletion of undernourished subjects alike show large gains in lean mass, with scant increase in fat. Androgens reduce fat to a degree but take away more lean mass. A low-energy diet for nonobese subjects removes more lean mass than it does fat, but for the obese it removes more fat than lean. Microgravity in orbit, like bed rest, adds a little fat and takes away a little of the lean, particularly from hard bone. Yet the near-weightless fetus and the infant in the crib grow splendidly. There is a great deal more to be learned along this road, the physiology of systems, here viewed at a considerable distance from the macromolecule.

ANNUAL INDEX

The following index lists all the authors and articles that appeared in SCIENTIFIC AMERICAN during 1987. Also indexed are "The Amateur Scientist" and "Computer Recreations."

AUTHORS

Abu-Mostafa, Yaser S., and Demetri Psaltis. OPTICAL NEURAL COMPUTERS; March, page 88.

Ada, Gordon L., and Sir Gustav Nossal. THE CLONAL-SELECTION THEORY; August, page 62.

Allen, Robert Day. THE MICROTUBULE AS AN INTRACELLULAR ENGINE; February, page 42.

Appenzeller, Tim, and Mortimer Mishkin. THE ANATOMY OF MEMORY; June, page 80.

Balzhiser, Richard E., and Kurt E. Yeager. COAL-FIRED POWER PLANTS FOR THE FUTURE; September, page 100.

Barrett, Spencer C. H. MIMICRY IN PLANTS; September, page 76.

Baruch, Jordan J., James Brian Quinn and Penny Cushman Paquette. TECHNOLOGY IN SERVICES; December, page 50.

Baylor, Denis A., and Julie L. Schnapf. HOW PHOTORECEPTOR CELLS RESPOND TO LIGHT; April, page 40.

Beck, Gregory, Gail S. Habicht and Jorge L. Benach. LYME DISEASE; July, page 78.

Benach, Jorge L., Gail S. Habicht and Gregory Beck. LYME DISEASE; July, page 78.

Bennett, Charles H. DEMONS, ENGINES AND THE SECOND LAW; November, page 108.

Bernstein, Barton J. THE BIRTH OF THE U.S. BIOLOGICAL-WARFARE PROGRAM; June, page 116.

Bharathan, Desikan, and Terry R. Penney. POWER FROM THE SEA; January, page 86.

Bloembergen, Nicolaas, and C. Kumar N. Patel. STRATEGIC DEFENSE AND DIRECTED-ENERGY WEAPONS; September, page 39.

Bonatti, Enrico. THE RIFTING OF CONTINENTS; March, page 96.

Bretscher, Mark S. HOW ANIMAL CELLS MOVE; December, page 72.

Brown, J. Larry. HUNGER IN THE U.S.; February, page 36.

Brown, Robert Hamilton, Torrence V. Johnson and Laurence A. Soderblom. THE MOONS OF URANUS; April, page 48.

Brownlee, Michael, Anthony Cerami and Helen Vlassara. GLUCOSE AND AGING; May, page 90.

Brownlee, Walter. H.M.S. *WARRIOR*; December, page 130.

Bunker, Bruce C., and Terry A. Michalske. THE FRACTURING OF GLASS; December, page 122.

Cerami, Anthony, Helen Vlassara and Michael Brownlee. GLUCOSE AND AGING; May, page 90.

Childress, James J., Horst Felbeck and George N. Somero. SYMBIOSIS IN THE DEEP SEA; May, page 114.

Chow, Marie, James M. Hogle and David J. Filman. THE STRUCTURE OF POLIOVIRUS; March, page 42.

Cohen, Leonard A. DIET AND CANCER; November, page 42.

Cole, Jonathan R., and Harriet Zuckerman. MARRIAGE, MOTHERHOOD AND RESEARCH PERFORMANCE IN SCIENCE; February, page 119.

Crews, David. COURTSHIP IN UNISEXUAL LIZARDS: A MODEL FOR BRAIN EVOLUTION; December, page 116.

Cuzzi, Jeffrey N., and Larry W. Esposito. THE RINGS OF URANUS; July, page 52.

Davis, Dan M., and Lynn R. Sykes. THE YIELDS OF SOVIET STRATEGIC WEAPONS; January, page 29.

Davis, William C. THE DETONATION OF EXPLOSIVES; May, page 106.

Dolan, Robert, and Harry Lins. BEACHES AND BARRIER ISLANDS; July, page 68.

Dressler, Alan. THE LARGE-SCALE STREAMING OF GALAXIES; September, page 46.

Dye, James L. ELECTRIDES; September, page 66.

Eastman, Lester F., and Mordehai Heiblum. BALLISTIC ELECTRONS IN SEMICONDUCTORS; February, page 102.

Erisman, Albert M., and Kenneth W. Neves. ADVANCED COMPUTING FOR MANUFACTURING; October, page 162.

Esposito, Larry W., and Jeffrey N. Cuzzi. THE RINGS OF URANUS; July, page 52.

Felbeck, Horst, James J. Childress and George N. Somero. SYMBIOSIS IN THE DEEP SEA; May, page 114.

Filman, David J., James M. Hogle and Marie Chow. THE STRUCTURE OF POLIOVIRUS; March, page 42.

Foley, James D. INTERFACES FOR ADVANCED COMPUTING; October, page 126.

Follett, R. F., and J. F. Power. MONOCULTURE; March, page 78.

Fox, Geoffrey C., and Paul C. Messina. ADVANCED COMPUTER ARCHITECTURES; October, page 66.

Francis, Peter, and Stephen Self. COLLAPSING VOLCANOES; June, page 90.

Frensley, William R. GALLIUM ARSENIDE TRANSISTORS; August, page 80.

Gallo, Robert C. THE AIDS VIRUS; January, page 46.

Gelernter, David. PROGRAMMING FOR ADVANCED COMPUTING; October, page 90.

Gildea, Patricia M., and George A. Miller. HOW CHILDREN LEARN WORDS; September, page 94.

Glantz, Michael H. DROUGHT IN AFRICA; June, page 34.

Gould, James L., and Peter Marler. LEARNING BY INSTINCT; January, page 74.

Greenberg, David A. MODELING TIDAL POWER; November, page 128.

Habicht, Gail S., Gregory Beck and Jorge L. Benach. LYME DISEASE; July, page 78.

Hamakawa, Yoshihiro. PHOTOVOLTAIC POWER; April, page 86.

Harkness, Robert P., and J. Craig Wheeler. HELIUM-RICH SUPERNOVAS; November, page 50.

Heiblum, Mordehai, and Lester F. Eastman. BALLISTIC ELECTRONS IN SEMICONDUCTORS; February, page 102.

Heinrich, Bernd. THERMOREGULATION IN WINTER MOTHS; March, page 104.

Hillis, W. Daniel. THE CONNECTION MACHINE; June, page 108.

Hirsch, Martin S., and Joan C. Kaplan. ANTIVIRAL THERAPY; April, page 76.

Hogle, James M., Marie Chow and David J. Filman. THE STRUCTURE OF POLIOVIRUS; March, page 42.

Hopfield, John J., and David W. Tank. COLLECTIVE COMPUTATION IN NEURONLIKE CIRCUITS; December, page 104.

Horton, Mark. THE SWAHILI CORRIDOR; September, page 86.

Howell, John M. EARLY FARMING IN NORTHWESTERN EUROPE; November, page 118.

Hut, Piet, and Gerald Jay Sussman. ADVANCED COMPUTING FOR SCIENCE; October, page 144.

Ingersoll, Andrew P. URANUS; January, page 38.

Jackson, Martin P. A., and Christopher J. Talbot. SALT TECTONICS; August, page 70.

Jeffries, Andrew D., Peter R. Saulson, Robert E. Spero and Michael E. Zucker. GRAVITATIONAL WAVE OBSERVATORIES; June, page 50.

Johnson, Kirk R., and C. Hans Nelson. WHALES AND WALRUSES AS TILLERS OF THE SEA FLOOR; February, page 112.

Johnson, Torrence V., Robert Hamilton Brown and Laurence A. Soderblom. THE MOONS OF URANUS; April, page 48.

Jones, Steven E., and Johann Rafelski.

COLD NUCLEAR FUSION; July, page 84.
Kahn, Robert E. NETWORKS FOR ADVANCED COMPUTING; October, page 136.
Kaplan, Joan C., and Martin S. Hirsch. ANTIVIRAL THERAPY; April, page 76.
Koch, Christof, and Tomaso Poggio. SYNAPSES THAT COMPUTE MOTION; May, page 46.
Krisch, Alan D. COLLISIONS BETWEEN SPINNING PROTONS; August, page 42.
Kryder, Mark H. DATA-STORAGE TECHNOLOGIES FOR ADVANCED COMPUTING; October, page 116.
Legge, Anthony J., and Peter A. Rowley-Conwy. GAZELLE KILLING IN STONE AGE SYRIA; August, page 88.
Letcher, Jr., John S., John K. Marshall, James C. Oliver III and Nils Salvesen. STARS & STRIPES; August, page 34.
Lins, Harry, and Robert Dolan. BEACHES AND BARRIER ISLANDS; July, page 68.
McMenamin, Mark A. S. THE EMERGENCE OF ANIMALS; April, page 94.
Marler, Peter, and James L. Gould. LEARNING BY INSTINCT; January, page 74.
Marrs, Barry L., and Douglas C. Youvan. MOLECULAR MECHANISMS OF PHOTOSYNTHESIS; June, page 42.
Marshall, John K., John S. Letcher, Jr., James C. Oliver III and Nils Salvesen. STARS & STRIPES; August, page 34.
Mathews, Grant J., and Victor E. Viola. THE COSMIC SYNTHESIS OF LITHIUM, BERYLLIUM AND BORON; May, page 38.
Mathews, Max V., and John R. Pierce. THE COMPUTER AS A MUSICAL INSTRUMENT; February, page 126.
Meindl, James D. CHIPS FOR ADVANCED COMPUTING; October, page 78.
Messina, Paul C., and Geoffrey C. Fox. ADVANCED COMPUTER ARCHITECTURES; October, page 66.
Metcalf, Harold J., and William D. Phillips. COOLING AND TRAPPING ATOMS; March, page 50.
Michalske, Terry A., and Bruce C. Bunker. THE FRACTURING OF GLASS; December, page 122.
Miller, George A., and Patricia M. Gildea. HOW CHILDREN LEARN WORDS; September, page 94.
Mishkin, Mortimer, and Tim Appenzeller. THE ANATOMY OF MEMORY; June, page 80.
Murray, Andrew W., and Jack W. Szostak. ARTIFICIAL CHROMOSOMES; November, page 62.
Nelson, C. Hans, and Kirk R. Johnson. WHALES AND WALRUSES AS TILLERS OF THE SEA FLOOR; February, page 112.
Neves, Kenneth W., and Albert M. Erisman. ADVANCED COMPUTING FOR MANUFACTURING; October, page 162.
Niklas, Karl J. AERODYNAMICS OF WIND POLLINATION; July, page 90.
Nossal, Sir Gustav, and Gordon L. Ada. THE CLONAL-SELECTION THEORY; August, page 62.
O'Brien, Stephen J. THE ANCESTRY OF THE GIANT PANDA; November, page 102.
Oliver III, James C., John S. Letcher, Jr., John K. Marshall and Nils Salvesen. STARS & STRIPES; August, page 34.
Ostro, Marc J. LIPOSOMES; January, page 102.
Paquette, Penny Cushman, James Brian Quinn and Jordan J. Baruch. TECHNOLOGY IN SERVICES; December, page 50.
Patel, C. Kumar N., and Nicolaas Bloembergen. STRATEGIC DEFENSE AND DIRECTED-ENERGY WEAPONS; September, page 39.
Patterson, David. THE CAUSES OF DOWN SYNDROME; August, page 52.
Peled, Abraham. THE NEXT COMPUTER REVOLUTION; October, page 56.
Penney, Terry R., and Desikan Bharathan. POWER FROM THE SEA; January, page 86.
Petersen, Erik Brinch, and T. Douglas Price. A MESOLITHIC CAMP IN DENMARK; March, page 112.
Phillips, William D., and Harold J. Metcalf. COOLING AND TRAPPING ATOMS; March, page 50.
Pierce, John R., and Max V. Mathews. THE COMPUTER AS A MUSICAL INSTRUMENT; February, page 126.
Poggio, Tomaso, and Christof Koch. SYNAPSES THAT COMPUTE MOTION; May, page 46.
Power, J. F., and R. F. Follett. MONOCULTURE; March, page 78.
Price, T. Douglas, and Erik Brinch Petersen. A MESOLITHIC CAMP IN DENMARK; March, page 112.
Psaltis, Demetri, and Yaser S. Abu-Mostafa. OPTICAL NEURAL COMPUTERS; March, page 88.
Quinn, James Brian, Jordan J. Baruch and Penny Cushman Paquette. TECHNOLOGY IN SERVICES; December, page 50.
Rafelski, Johann, and Steven E. Jones. COLD NUCLEAR FUSION; July, page 84.
Rennels, Glenn D., and Edward H. Shortliffe. ADVANCED COMPUTING FOR MEDICINE; October, page 154.
Rowley-Conwy, Peter A., and Anthony J. Legge. GAZELLE KILLING IN STONE AGE SYRIA; August, page 88.
Runcorn, S. K. THE MOON'S ANCIENT MAGNETISM; December, page 60.
Ryder, Michael L. THE EVOLUTION OF THE FLEECE; January, page 112.
Salvesen, Nils, John S. Letcher, Jr., John K. Marshall and James C. Oliver III. STARS & STRIPES; August, page 34.
Sander, Leonard M. FRACTAL GROWTH; January, page 94.
Saulson, Peter R., Andrew D. Jeffries, Robert E. Spero and Michael E. Zucker. GRAVITATIONAL WAVE OBSERVATORIES; June, page 50.
Schnapf, Julie L., and Denis A. Baylor. HOW PHOTORECEPTOR CELLS RESPOND TO LIGHT; April, page 40.
Schneider, Stephen H. CLIMATE MODELING; May, page 72.
Self, Stephen, and Peter Francis. COLLAPSING VOLCANOES; June, page 90.
Shaham, Jacob. THE OLDEST PULSARS IN THE UNIVERSE; February, page 50.
Shaw, Robert W. AIR POLLUTION BY PARTICLES; August, page 96.
Shen, Sinyan. ACOUSTICS OF ANCIENT CHINESE BELLS; April, page 104.
Shortliffe, Edward H., and Glenn D. Rennels. ADVANCED COMPUTING FOR MEDICINE; October, page 154.
Soderblom, Laurence A., Torrence V. Johnson and Robert Hamilton Brown. THE MOONS OF URANUS; April, page 48.
Somero, George N., James J. Childress and Horst Felbeck. SYMBIOSIS IN THE DEEP SEA; May, page 114.
Spero, Robert E., Andrew D. Jeffries, Peter R. Saulson and Michael E. Zucker. GRAVITATIONAL WAVE OBSERVATORIES; June, page 50.
Stahl, Franklin W. GENETIC RECOMBINATION; February, page 90.
Stryer, Lubert. THE MOLECULES OF VISUAL EXCITATION; July, page 42.
Sussman, Gerald Jay, and Piet Hut. ADVANCED COMPUTING FOR SCIENCE; October, page 144.
Sykes, Lynn R., and Dan M. Davis. THE YIELDS OF SOVIET STRATEGIC WEAPONS; January, page 29.
Szekely, Julian. CAN ADVANCED TECHNOLOGY SAVE THE U.S. STEEL INDUSTRY?; July, page 34.
Szostak, Jack W., and Andrew W. Murray. ARTIFICIAL CHROMOSOMES; November, page 62.
Talbot, Christopher J., and Martin P. A. Jackson. SALT TECTONICS; August, page 70.
Tank, David W., and John J. Hopfield. COLLECTIVE COMPUTATION IN NEURONLIKE CIRCUITS; December, page 104.
Taylor, Theodore B. THIRD-GENERATION NUCLEAR WEAPONS; April, page 30.
Thurow, Lester C. A SURGE IN INEQUALITY; May, page 30.
Toth, Nicholas. THE FIRST TECHNOLOGY; April, page 112.
Van Beek, Gus W. ARCHES AND VAULTS IN THE ANCIENT NEAR EAST; July, page 96.
Varmus, Harold. REVERSE TRANSCRIPTION; September, page 56.
Vause, Chester A., and James S. Walk-

er. REAPPEARING PHASES; May, page 98.
Vellutino, Frank R. DYSLEXIA; March, page 34.
Vilenkin, Alexander. COSMIC STRINGS; December, page 94.
Viola, Victor E., and Grant J. Mathews. THE COSMIC SYNTHESIS OF LITHIUM, BERYLLIUM AND BORON; May, page 38.
Vlassara, Helen, Anthony Cerami and Michael Brownlee. GLUCOSE AND AGING; May, page 90.
Walker, James S., and Chester A. Vause. REAPPEARING PHASES; May, page 98.
Wheeler, J. Craig, and Robert P. Harkness. HELIUM-RICH SUPERNOVAS; November, page 50.
Winick, Herman. SYNCHROTRON RADIATION; November, page 88.
Yeager, Kurt E., and Richard E. Balzhiser. COAL-FIRED POWER PLANTS FOR THE FUTURE; September, page 100.
Youvan, Douglas C., and Barry L. Marrs. MOLECULAR MECHANISMS OF PHOTOSYNTHESIS; June, page 42.
Zapol, Warren M. DIVING ADAPTATIONS OF THE WEDDELL SEAL; June, page 100.
Zucker, Michael E., Andrew D. Jeffries, Peter R. Saulson and Robert E. Spero. GRAVITATIONAL WAVE OBSERVATORIES; June, page 50.
Zuckerman, Harriet, and Jonathan R. Cole. MARRIAGE, MOTHERHOOD AND RESEARCH PERFORMANCE IN SCIENCE; February, page 119.

ARTICLES

AIDS VIRUS, THE, by Robert C. Gallo; January, page 46.
ANIMALS, THE EMERGENCE OF, by Mark A. S. McMenamin; April, page 94.
ANTIVIRAL THERAPY, by Martin S. Hirsch and Joan C. Kaplan; April, page 76.
ARCHES AND VAULTS IN THE ANCIENT NEAR EAST, by Gus W. Van Beek; July, page 96.
ATOMS, COOLING AND TRAPPING, by William D. Phillips and Harold J. Metcalf; March, page 50.
BALLISTIC ELECTRONS IN SEMICONDUCTORS, by Mordehai Heiblum and Lester F. Eastman; February, page 102.
BEACHES AND BARRIER ISLANDS, by Robert Dolan and Harry Lins; July, page 68.
BIOLOGICAL-WARFARE PROGRAM, THE BIRTH OF THE U.S., by Barton J. Bernstein;June, page 116.
CANCER, DIET AND, by Leonard A. Cohen; November, page 42.
CELLS MOVE, HOW ANIMAL, by Mark S. Bretscher; December, page 72.
CHILDREN LEARN WORDS, HOW, by George A. Miller and Patricia M. Gildea; September, page 94.
CHINESE BELLS, ACOUSTICS OF ANCIENT, by Sinyan Shen; April, page 104.
CHROMOSOMES, ARTIFICIAL, by Andrew W. Murray and Jack W. Szostak; November, page 62.
CLIMATE MODELING, by Stephen H. Schneider; May, page 72.
CLONAL-SELECTION THEORY, THE, by Gordon L. Ada and Sir Gustav Nossal; August, page 62.
COAL-FIRED POWER PLANTS FOR THE FUTURE, by Richard E. Balzhiser and Kurt E. Yeager; September, page 100.
COMPUTER ARCHITECTURES, ADVANCED, by Geoffrey C. Fox and Paul C. Messina; October, page 66.
COMPUTER AS A MUSICAL INSTRUMENT, THE, by Max V. Mathews and John R. Pierce; February, page 126.
COMPUTER REVOLUTION, THE NEXT, by Abraham Peled; October, page 56.
COMPUTING, CHIPS FOR ADVANCED, by James D. Meindl; October, page 78.
COMPUTING, DATA-STORAGE TECHNOLOGIES FOR ADVANCED, by Mark H. Kryder; October, page 116.
COMPUTING FOR MANUFACTURING, ADVANCED, by Albert M. Erisman and Kenneth W. Neves; October, page 162.
COMPUTING FOR MEDICINE, ADVANCED, by Glenn D. Rennels and Edward H. Shortliffe; October, page 154.
COMPUTING FOR SCIENCE, ADVANCED, by Piet Hut and Gerald Jay Sussman; October, page 144.
COMPUTING, INTERFACES FOR ADVANCED, by James D. Foley; October, page 126.
COMPUTING, NETWORKS FOR ADVANCED, by Robert E. Kahn; October, page 136.
COMPUTING, PROGRAMMING FOR ADVANCED, by David Gelernter; October, page 90.
CONNECTION MACHINE, THE, by W. Daniel Hillis; June, page 108.
COSMIC STRINGS, by Alexander Vilenkin; December, page 94.
COURTSHIP IN UNISEXUAL LIZARDS: A MODEL FOR BRAIN EVOLUTION, by David Crews; December, page 116.
DEMONS, ENGINES AND THE SECOND LAW, by Charles H. Bennett; November, page 108.
DOWN SYNDROME, THE CAUSES OF, by David Patterson; August, page 52.
DROUGHT IN AFRICA, by Michael H. Glantz; June, page 34.
DYSLEXIA, by Frank R. Vellutino; March, page 34.
ELECTRIDES, by James L. Dye; September, page 66.
EXPLOSIVES, THE DETONATION OF, by William C. Davis; May, page 106.
FARMING IN NORTHWESTERN EUROPE, EARLY, by John M. Howell; November, page 118.
FLEECE, THE EVOLUTION OF THE, by Michael L. Ryder; January, page 112.
FRACTAL GROWTH, by Leonard M. Sander; January, page 94.
GALAXIES, THE LARGE-SCALE STREAMING OF, by Alan Dressler; September, page 46.
GALLIUM ARSENIDE TRANSISTORS, by William R. Frensley; August, page 80.
GAZELLE KILLING IN STONE AGE SYRIA, by Anthony J. Legge and Peter A. Rowley-Conwy; August, page 88.
GLASS, THE FRACTURING OF, by Terry A. Michalske and Bruce C. Bunker; December, page 122.
GLUCOSE AND AGING, by Anthony Cerami, Helen Vlassara and Michael Brownlee; May, page 90.
GRAVITATIONAL WAVE OBSERVATORIES, by Andrew D. Jeffries, Peter R. Saulson, Robert E. Spero and Michael E. Zucker; June, page 50.
HUNGER IN THE U.S., by J. Larry Brown; February, page 36.
INEQUALITY, A SURGE IN, by Lester C. Thurow; May, page 30.
LEARNING BY INSTINCT, by James L. Gould and Peter Marler; January, page 74.
LIPOSOMES, by Marc J. Ostro; January, page 102.
LYME DISEASE, by Gail S. Habicht, Gregory Beck and Jorge L. Benach; July, page 78.
MEMORY, THE ANATOMY OF, by Mortimer Mishkin and Tim Appenzeller; June, page 80.
MESOLITHIC CAMP IN DENMARK, A, by T. Douglas Price and Erik Brinch Petersen; March, page 112.
MICROTUBULE AS AN INTRACELLULAR ENGINE, THE, by Robert Day Allen; February, page 42.
MIMICRY IN PLANTS, by Spencer C. H. Barrett; September, page 76.
MONOCULTURE, by J. F. Power and R. F. Follett; March, page 78.
MOON'S ANCIENT MAGNETISM, THE, by S. K. Runcorn; December, page 60.
NEURAL COMPUTERS, OPTICAL, by Yaser S. Abu-Mostafa and Demetri Psaltis; March, page 88.
NEURONLIKE CIRCUITS, COLLECTIVE COMPUTATION IN, by David W. Tank and John J. Hopfield; December, page 104.
NUCLEAR FUSION, COLD, by Johann Rafelski and Steven E. Jones; July, page 84.
NUCLEAR WEAPONS, THIRD-GENERATION, by Theodore B. Taylor; April, page 30.
PANDA, THE ANCESTRY OF THE GIANT, by Stephen J. O'Brien; November, page 102.
PHOTORECEPTOR CELLS RESPOND TO

LIGHT, HOW, by Julie L. Schnapf and Denis A. Baylor; April, page 40.
PHOTOSYNTHESIS, MOLECULAR MECHANISMS OF, by Douglas C. Youvan and Barry L. Marrs; June, page 42.
PHOTOVOLTAIC POWER, by Yoshihiro Hamakawa; April, page 86.
POLIOVIRUS, THE STRUCTURE OF, by James M. Hogle, Marie Chow and David J. Filman; March, page 42.
POLLINATION, AERODYNAMICS OF WIND, by Karl J. Niklas; July, page 90.
POLLUTION BY PARTICLES, AIR, by Robert W. Shaw; August, page 96.
POWER FROM THE SEA, by Terry R. Penney and Desikan Bharathan; January, page 86.
PROTONS, COLLISIONS BETWEEN SPINNING, by Alan D. Krisch; August, page 42.
PULSARS IN THE UNIVERSE, THE OLDEST, by Jacob Shaham; February, page 50.
REAPPEARING PHASES, by James S. Walker and Chester A. Vause; May, page 98.
RECOMBINATION, GENETIC, by Franklin W. Stahl; February, page 90.
RESEARCH PERFORMANCE IN SCIENCE, MARRIAGE, MOTHERHOOD AND, by Jonathan R. Cole and Harriet Zuckerman; February, page 119.
REVERSE TRANSCRIPTION, by Harold Varmus; September, page 56.
RIFTING OF CONTINENTS, THE, by Enrico Bonatti; March, page 96.
SALT TECTONICS, by Christopher J. Talbot and Martin P. A. Jackson; August, page 70.
SEA FLOOR, WHALES AND WALRUSES AS TILLERS OF THE, by C. Hans Nelson and Kirk R. Johnson; February, page 112.
SEAL, DIVING ADAPTATIONS OF THE WEDDELL, by Warren M. Zapol; June, page 100.
SERVICES, TECHNOLOGY IN, by James Brian Quinn, Jordan J. Baruch and Penny Cushman Paquette; December, page 50.
STARS & STRIPES, by John S. Letcher, Jr., John K. Marshall, James C. Oliver III and Nils Salvesen; August, page 34.
STEEL INDUSTRY?, CAN ADVANCED TECHNOLOGY SAVE THE U.S., by Julian Szekely; July, page 34.
STRATEGIC DEFENSE AND DIRECTED-ENERGY WEAPONS, by C. Kumar N. Patel and Nicolaas Bloembergen; September, page 39.
SUPERNOVAS, HELIUM-RICH, by J. Craig Wheeler and Robert P. Harkness; November, page 50.
SWAHILI CORRIDOR, THE, by Mark Horton; September, page 86.
SYMBIOSIS IN THE DEEP SEA, by James J. Childress, Horst Felbeck and George N. Somero; May, page 114.
SYNAPSES THAT COMPUTE MOTION, by Tomaso Poggio and Christof Koch; May, page 46.
SYNCHROTRON RADIATION, by Herman Winick; November, page 88.
SYNTHESIS OF LITHIUM, BERYLLIUM AND BORON, THE COSMIC, by Victor E. Viola and Grant J. Mathews; May, page 38.
TECHNOLOGY, THE FIRST, by Nicholas Toth; April, page 112.
THERMOREGULATION IN WINTER MOTHS, by Bernd Heinrich; March, page 104.
TIDAL POWER, MODELING, by David A. Greenberg; November, page 128.
URANUS, by Andrew P. Ingersoll; January, page 38.
URANUS, THE MOONS OF, by Torrence V. Johnson, Robert Hamilton Brown and Laurence A. Soderblom; April, page 48.
URANUS, THE RINGS OF, by Jeffrey N. Cuzzi and Larry W. Esposito; July, page 52.
VISUAL EXCITATION, THE MOLECULES OF, by Lubert Stryer; July, page 42.
VOLCANOES, COLLAPSING, by Peter Francis and Stephen Self; June, page 90.
WARRIOR, H.M.S., by Walter Brownlee; December, page 130.
YIELDS OF SOVIET STRATEGIC WEAPONS, THE, by Lynn R. Sykes and Dan M. Davis; January, page 29.

THE AMATEUR SCIENTIST

Barometer that works with water in place of mercury, Making a; April, page 122.
Bead arrays, Sticky threadlike substances that tend to draw themselves out into; September, page 108.
Cheshire cat's odd vanishing act, Concerning disappearances, including the; May, page 122.
Color pattern of a soap film, Music and ammonia vapor excite the; August, page 104.
Disappearances, including the Cheshire cat's odd vanishing act, Concerning; May, page 122.
Distance to the sun by observing the trail of a meteor, Calculating the; March, page 122.
Fluid flows faster when the tube is pinched, Why a; July, page 104.
Hele-Shaw cell, Fluid interfaces, including fractal flows, can be studied in a; November, page 134.
Microwave oven's rapid cooking action is disclosed, The secret of a; February, page 134.
Puzzles in two and three dimensions, and ways to simplify their solution; June, page 122.
Reflections from a water surface display some curious properties; January, page 120.
Rubik's Magic, a new puzzle that provides a study in permutation operators, Now there is; October, page 170.
Soap film, Music and ammonia vapor excite the color pattern of a; August, page 104.
Tape is peeled off a surface, How to capture on film the faint glow emitted when sticky; December, page 138.

COMPUTER RECREATIONS

"After MAD": a computer game of nuclear strategy that ends in a Prisoner's Dilemma; October, page 174.
Algopuzzles: wherein trains of thought follow algorithmic tracks to solutions; June, page 128.
Algorithms into programs, Simple special effects illustrate the art of converting; December, page 142.
Braitenberg memoirs: vehicles for probing behavior roam a dark plain marked by lights; March, page 16.
Bulls, bears and programs in the pit, Of; May, page 16.
Chaos, Probing the strange attractions of; July, page 108.
Computer game of nuclear strategy that ends in a Prisoner's Dilemma, "After MAD": a; October, page 174.
Computer party, Diverse personalities search for social equilibrium at a; September, page 112.
Computing is music to the ears of some, The sound of; April, page 14.
Core War tournament, A program called MICE nibbles its way to victory at the first; January, page 14.
Julia, Beauty and profundity: the Mandelbrot set and a flock of its cousins called; November, page 140.
Life acquires some successors in three dimensions, The game; February, page 16.
Mandelbrot set and a flock of its cousins called Julia, Beauty and profundity: the; November, page 140.
Music to the ears of some, The sound of computing is; April, page 14.
Special effects illustrate the art of converting algorithms into programs, Simple; December, page 142.
Strange attractions of chaos, Probing the; July, page 108.
Trains of thought follow algorithmic tracks to solutions, Algopuzzles: wherein; June, page 128.
Vehicles for probing behavior roam a dark plain marked by lights, Braitenberg memoirs; March, page 16.
Word ladders and a tower of Babel lead to computational heights defying assault; August, page 108.

BIBLIOGRAPHY

Readers interested in further exploration of the subjects covered by the articles in this issue may find the following lists of publications helpful.

THE ANTARCTIC OZONE HOLE

NIMBUS 7 SATELLITE MEASUREMENTS OF THE SPRINGTIME ANTARCTIC OZONE DECREASE. R. S. Stolarski, A. J. Krueger, M. R. Schoeberl, R. D. McPeters, P. A. Newman and J. C. Alpert in *Nature,* Vol. 322, No. 6082, pages 808–811; August 28, 1986.

ANTARCTIC OZONE DEPLETION. In *Geophysical Research Letters,* Vol. 13, No. 12; November Supplement, 1986.

STRATOSPHERIC TRACE GASES IN THE SPRING 1986 ANTARCTIC ATMOSPHERE. C. B. Farmer, G. C. Toon, P. W. Schaper, J.-F. Blavier and L. L. Lowes in *Nature,* Vol. 329, No. 6135, pages 126–130; September 10, 1987.

HOW KILLER CELLS KILL

MECHANISM OF LYMPHOCYTE-MEDIATED CYTOTOXICITY. Pierre A. Henkart in *Annual Review of Immunology,* Vol. 3, pages 31–58; 1985.

MOLECULAR MECHANISM OF LYMPHOCYTE MEDIATED TUMOR LYSIS. Eckhard R. Podack in *Immunology Today,* Vol. 6, pages 21–27; 1985.

PURIFICATION AND CHARACTERIZATION OF A CYTOLYTIC PORE-FORMING PROTEIN FROM GRANULES OF CLONED LYMPHOCYTES WITH NATURAL KILLER ACTIVITY. John Ding-E Young, Hans Hengartner, Eckhard R. Podack and Zanvil A. Cohn in *Cell,* Vol. 44, No. 6, pages 849–859; March 28, 1986.

A CALCIUM- AND PERFORIN-INDEPENDENT PATHWAY OF KILLING MEDIATED BY MURINE CYTOLYTIC LYMPHOCYTES. John Ding-E Young, William R. Clark, Chau-Ching Liu and Zanvil A. Cohn in *The Journal of Experimental Medicine,* Vol. 166, No. 6, pages 1894–1899; December, 1987.

THE REALITY OF THE QUANTUM WORLD

QUANTUM THEORY AND MEASUREMENT. Edited by John Archibald Wheeler and Wojciech Hubert Zurek. Princeton University Press, 1983.

FOUNDATIONS OF QUANTUM MECHANICS IN THE LIGHT OF NEW TECHNOLOGY. Edited by S. Kamefuchi, H. Ezawa, Y. Murayama, M. Namiki, S. Nomura, Y. Ohnuki and T. Yajima. Physical Society of Japan, 1984.

THE QUANTUM WORLD. J. C. Polkinghorne. Princeton University Press, 1985.

QUANTUM PHYSICS: ILLUSION OR REALITY? Alastair I. M. Rae. Cambridge University Press, 1986.

THE VERY-LONG-BASELINE ARRAY

THE VERY LARGE ARRAY: DESIGN AND PERFORMANCE OF A MODERN SYNTHESIS RADIO TELESCOPE. Peter J. Napier, A. Richard Thompson and Ronald D. Ekers in *Proceedings of the IEEE,* Vol. 71, No. 11, pages 1295–1320; November, 1983.

INTERFEROMETRY AND SYNTHESIS IN RADIO ASTRONOMY. A. Richard Thompson, James Moran and G. W. Swenson, Jr. John Wiley & Sons, Inc., 1986.

THE INVISIBLE UNIVERSE REVEALED: THE STORY OF RADIO ASTRONOMY. G. L. Verschuur. Springer-Verlag, 1986.

INTERTIDAL FISHES

RECENT STUDIES ON THE BIOLOGY OF INTERTIDAL FISHES. R. N. Gibson in *Oceanography and Marine Biology: An Annual Review,* Vol. 20, pages 363–414; 1982.

INTERTIDAL FISHES: LIFE IN A FLUCTUATING ENVIRONMENT. Robin N. Gibson in *The Behavior of Teleost Fishes,* edited by Tony J. Pitcher. The Johns Hopkins University Press, 1986.

HERBIVORE RESPONSES TO A SEASONALLY FLUCTUATING FOOD SUPPLY: GROWTH POTENTIAL OF TWO TEMPERATE INTERTIDAL FISHES BASED ON THE PROTEIN AND ENERGY ASSIMILATED FROM THEIR MACROALGAL DIETS. Michael H. Horn, Margaret A. Neighbors and Steven N. Murray in *Journal of Experimental Marine Biology and Ecology,* Vol. 103, No. 1, 2, 3, pages 217–234; December 16, 1986.

THE NOT-SO-RARE EARTHS

LANTHANOIDS AS GEOCHEMICAL PROBE AND PROBLEMS IN LANTHANOID GEOCHEMISTRY—DISTRIBUTION AND BEHAVIOUR OF LANTHANOIDS IN NONMAGMATIC PHASES. Peter Möller in *Systematics and the Property of the Lanthanides,* edited by S. P. Sinha. D. Reidel Publishing Company, 1983.

RARE EARTH ELEMENT GEOCHEMISTRY. Edited by Paul Henderson. Elsevier North-Holland Publishing Co., 1984.

THE CONTINENTAL CRUST: ITS COMPOSITION AND EVOLUTION. Stuart Ross Taylor and Scott M. McLennan. Blackwell Scientific Publications, 1985.

ART, ILLUSION AND THE VISUAL SYSTEM

FUNCTIONAL ARCHITECTURE OF MACAQUE MONKEY VISUAL CORTEX. D. H. Hubel and T. N. Wiesel in *Proceedings of the Royal Society of London, Series B,* Vol. 198, No. 1130, pages 1–59; July, 1977.

ANATOMY AND PHYSIOLOGY OF A COLOR SYSTEM IN THE PRIMATE VISUAL CORTEX. Margaret S. Livingstone and David H. Hubel in *The Journal of Neuroscience,* Vol. 4, No. 1, pages 309–356; January, 1984.

EYE, BRAIN, AND VISION. David H. Hubel. Scientific American Library, 1988.

THE TRANSFORMER

A LIFE OF GEORGE WESTINGHOUSE. Henry G. Prout. Charles Scribner's Sons, 1922.

THE RISE OF THE ELECTRICAL INDUSTRY DURING THE NINETEENTH CENTURY. Malcolm MacLaren. Princeton University Press, 1943.

HISTORY OF SOFT MAGNETIC MATERIALS. J. H. Bechtold and G. W. Wiener in *The Metallurgical Society Conferences: Vol. 27,* edited by Cyril Stanley Smith. Gordon & Breach Science Publishers, 1965.

THE AMATEUR SCIENTIST

SUN FIGURE: AN ILLUSORY DIFFUSE CONTOUR RESULTING FROM AN ARRANGEMENT OF DOTS. John M. Kennedy in *Perception,* Vol. 5, pages 479–481; 1976.

THE EFFECTS OF PERCEPTUAL SET ON THE SHAPE AND APPARENT DEPTH OF SUBJECTIVE CONTOURS. Stanley Coren, Clare Porac and Leonard H. Theodor in *Perception & Psychophysics,* Vol. 39, No. 5, pages 327–333; 1986.

COMPUTER RECREATIONS

MOLECULAR ENGINEERING: AN APPROACH TO THE DEVELOPMENT OF GENERAL CAPABILITIES FOR MOLECULAR MANIPULATION. K. Eric Drexler in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 78, No. 9, pages 5275–5278; September, 1981.

ENGINES OF CREATION. K. Eric Drexler. Anchor Press, 1986.


Street smart.
Doing business in Tokyo is a lot like catching a cab there: you've got to be in the right place at the right time. United helps by flying you to Tokyo from more U.S. cities than any other airline.
Along the way, you'll enjoy Friendly Skies service that provides the best in international travel. Including, in First Class, sleeper seats; and, on the ground, our exclusive Concierge Service.
It's the smartest way to go. United. Rededicated to giving you the service you deserve. Come fly the friendly skies.
UNITED
AIRLINES
TOKYO • OSAKA • HONG KONG • SEOUL • TAIPEI • SYDNEY • MELBOURNE • BEIJING • SHANGHAI • AUCKLAND • SINGAPORE • MANILA • BANGKOK
J&P
HOTEL LA PLAZA
白波
日月東
50