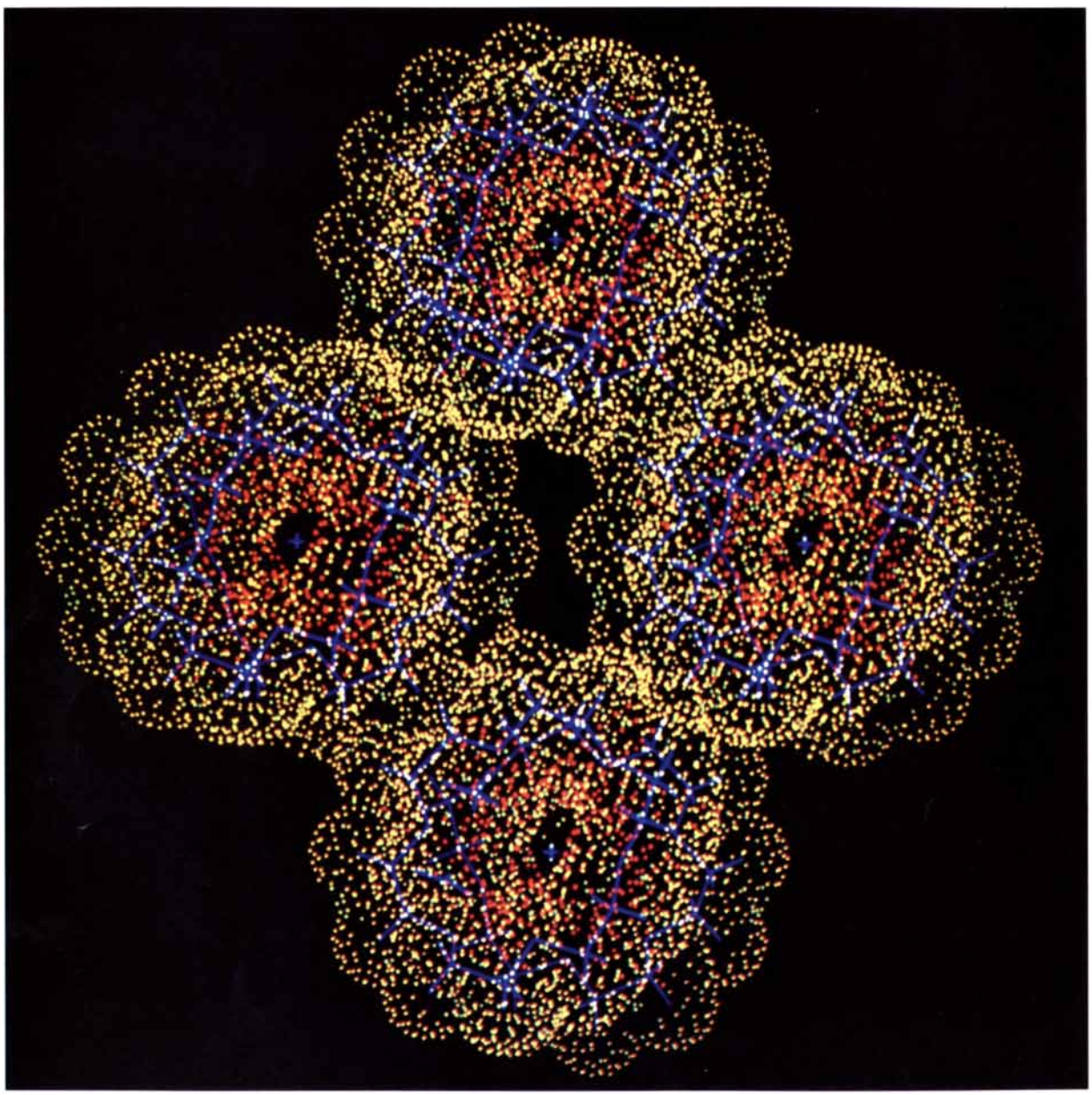
SEPTEMBER 1987
$2.50

SCIENTIFIC AMERICAN

Are galaxies streaming toward a distant Great Attractor?
By age 17 a human being may know 80,000 words; how does she do it?
To survive some plants play a devious game: they imitate insects.



A New Class of Crystalline Materials *called electrides shows interplay among electrons and could promise new electronic and optical devices.*

© 1987 SCIENTIFIC AMERICAN, INC

CAUTION: USE OF THIS PRODUCT MAY RESULT IN CHRONIC DISILLUSIONMENT WITH THE REST OF THE WORLD.

The new L6 luxury sports coupe from BMW is designed to induce a most gratifying form of selective amnesia in anyone who enters it.

It enables you to temporarily forget that this is less than a perfect world.

Because in a perfect world, for example, attention to detail and reverence for craftsmanship are paramount virtues. And in the BMW L6, it takes 100 artisans over 150 hours just to hand-sew and sculpt the supple, all-leather interior.

In a perfect world, intelligent, functional design is the norm. And in the BMW L6, the driver sits behind a curved, biomechanically-engineered instrument panel, with all vital controls and ancillary instruments within easy reach and view.

In fact, the BMW L6 is a refinement of a car that the editors of Road & Track likened to "stepping into the next century. Everything is high-tech, ergonomic, non-reflecting, purposeful. And ready to be turned loose."

© 1987 BMW of North America, Inc. The BMW trademark and logo are registered.

© 1987 SCIENTIFIC AMERICAN, INC

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

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

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

THE ULTIMATE DRIVING MACHINE.

© 1987 SCIENTIFIC AMERICAN, INC

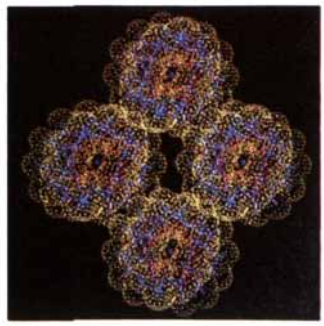
SCIENTIFIC AMERICAN

September 1987 Volume 257 Number 3

© 1987 SCIENTIFIC AMERICAN, INC

DEPARTMENTS

© 1987 SCIENTIFIC AMERICAN, INC



THE COVER image depicts the crystal structure of an electride: a material in which lone electrons occupy sites normally filled by negative ions (see "Electrides," by James L. Dye, page 66). Dots define the surfaces of atoms and blue skeletons represent carbon chains. Each of the four round clusters is a "sandwich" in which two neutral molecules enclose a positively charged cesium atom. Between the clusters there is a channel connecting cavities that hold trapped electrons.

THE ILLUSTRATIONS

Cover photograph by James Kilkelly

Page	Source
40	Gabor Kiss
41	White Sands Missile Test Range
42	Jon Brenneis
43	Lawrence Livermore National Laboratory
44	Gabor Kiss
45	Jon Brenneis
47	Ofer Lahav, Institute of Astronomy, University of Cambridge
48–49	Hank Iken
50	David T. Wilkinson, Princeton University
51–54	Hank Iken
57	Greg Monokian and Jef D. Boeke, Johns Hopkins University School of Medicine
58–64	Dana Burns
67	James Kilkelly
68	Andrew Christie
69–70	James Kilkelly
71–74	Andrew Christie
75	James Kilkelly
77	David Thompson, Oxford Scientific Films
78–79	Tom Prentiss
80	G. I. Bernard, Oxford Scientific Films
81	Breck P. Kent, Earth Scenes
82–83	Spencer C. H. Barrett
87	© Trustees of the Victoria and Albert Museum
88	Hank Iken
89–90	Patricia J. Wynne
91	M. C. Horton
92–93	Ashmolean Museum of Art and Archaeology, Oxford
95	Quesada/Burke
96	James Kilkelly
97	Quesada/Burke
98	George A. Miller
99	Edward Bell
101	Electric Power Research Institute
102–106	George V. Kelvin
108	Michael Goodman
109	C. C. Lockwood, Animals Animals (*top*); John Gosline, University of British Columbia (*bottom*)
110–111	Michael Goodman
112–113	Edward Bell
114	Edward Bell (*top*), Andrew Christie (*bottom*)

SCIENTIFIC AMERICAN

Established 1845

EDITOR Jonathan Piel

BOARD OF EDITORS Armand Schwab, Jr. (Managing Editor), Timothy Appenzeller (Associate Editor), Timothy M. Beardsley (Washington), John M. Benditt, Laurie Burnham, David L. Cooke, Jr., Ari W. Epstein, Gregory R. Greenwell, John Horgan ("Science and the Citizen"), Robert Kunzig, Philip Morrison (Book Editor), James T. Rogers, Ricki L. Rusting, Karen Wright

ART DEPARTMENT Samuel L. Howard (Art Director), Steven R. Black (Associate Art Director), Edward Bell (Assistant Art Director), Johnny Johnson

PRODUCTION DEPARTMENT Richard Sasso (Director of Production), Carol Eisler and Leo J. Petruzzi (Assistants to the Production Director), Carol Hansen (Electronic Composition Manager), Carol Albert, Nancy Mongelli, Jody Seward, William Sherman, Julio E. Xavier

COPY DEPARTMENT Sally Porter Jenks (Copy Chief), Debra Q. Bennett, M. Knight, Dorothy R. Patterson

PUBLISHER Harry Myers
ADVERTISING DIRECTOR Peter B. Kennedy
CIRCULATION MANAGER William H. Yokel
DIRECTOR OF OPERATIONS R. Vincent Barger

PRESIDENT Georg-Dieter von Holtzbrinck
EXECUTIVE COMMITTEE Claus-Gerhard Firchow, Executive Vice-President; Harry Myers, Senior Vice-President; R. Vincent Barger, Vice-President, Finance; Linda Chaput, Vice-President; Jonathan Piel, Vice-President; Carol Snow, Vice-President

CHAIRMAN OF THE BOARD Gerard Piel

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.

© 1987 SCIENTIFIC AMERICAN, INC

frustrated	aggravated	sick
impatient	confused	troubled
upset	perturbed	tired
tedious	overwhelmed	miffed
dizzy	defeated	agitated
perplexed	stupid	wrecked
crazy	annoyed	moronic
dumb	irate	pained
outraged	foiled	thwarted

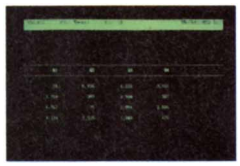
Don't hold back now. How do you really feel about working with columns?

Columns. The black hole of word processing technology.

Try to move a column with your present word processor. Good luck. Try to add or delete a column. Sorry.

Lotus Manuscript™ can change how you feel about working with columns. Really. Because Lotus Manuscript thinks vertically as well as horizontally.

Consider this. On ordinary word processors, when you create a table, or put any information in columns, you work horizontally, using tabs to create a vertical effect.



Inserting a column with your present word processor would be a true test of patience, right?

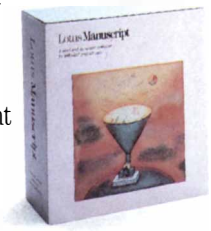
It's simple with Manuscript, because Manuscript thinks vertically as well as horizontally.

With our sophisticated table editor, you can actually create each column independent of one another. You can change information, even delete, move or insert columns, without the usual litany of expletives deleted.

And it's not just limited to columns you create yourself. Take tables from Lotus® 1-2-3® or Symphony,® for instance. When you import that tabular information into Lotus Manuscript, it recognizes it as columns, so you can change, delete, move, insert or format this material as well. You can even dress up your tables with a variety of boxes and borders.

As you can see, Lotus Manuscript is not just another word processor. It's a complete document creation system, with more impressive features than we could ever go into here.

Lotus Manuscript is designed to work on most IBM® PCs and compatibles.* Its familiar 1-2-3 interface makes it easy to use. And our Manuscript evaluation kit makes it easy to try. For $10.00, you'll get a presentation disk, working software, and a tutorial manual. To get your evaluation kit, call 1-800-345-1043, ask for lot BS-1450. Or, for more information, see your Authorized Lotus Dealer, or write Lotus Development Corp., 90 Annex, Atlanta, GA 30390-03070.



Lotus Manuscript™

© 1987 Lotus Development Corporation. Lotus, 1-2-3, and Symphony are registered trademarks and Lotus Manuscript is a trademark of Lotus Development Corporation. IBM is a registered trademark of International Business Machines. *Manuscript and Manuscript Evaluation Kit require 512K and a hard disk.

© 1987 SCIENTIFIC AMERICAN, INC

Northwest covers

Osaka

Tokyo

Hong Kong

Seoul

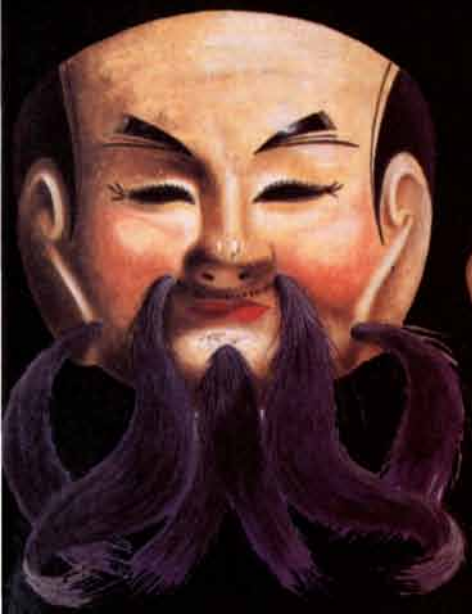
Bangkok

Guam

© 1987 Northwest Airlines, Inc.

© 1987 SCIENTIFIC AMERICAN, INC

the face of Asia.



Taipei

Okinawa

Manila

Shanghai

No matter how you look at it, Northwest has Asia covered. With convenient service from over 200 U.S. cities.

Including nonstops from New York, Chicago, Seattle, San Francisco, Los Angeles, Detroit and Honolulu.

And we're the only U.S. airline to give you the comforts of an all-747 transpacific fleet. The luxuries of Regal Imperial service. The experience of 40 years of flying to Asia. And the rewards of our WORLDPERKS℠ frequent flyer program, which gives you the fastest free trip of any major airline.

These are just some of the reasons why we're America's number one choice to Asia.

So if you're looking for a familiar face in the Far East, call your travel agent or call Northwest at (800) 447-4747 for international reservations.

Look to us.

NORTHWEST

© 1987 SCIENTIFIC AMERICAN, INC

LETTERS

To the Editors:

In reference to "Dyslexia," by Frank R. Vellutino [SCIENTIFIC AMERICAN, March], we should like to point out that there are many reading disorders, all coming under the name of dyslexia, and that mirror writing is the symptom of only one of them.

My daughter Arwen had severe reading problems in elementary school but never exhibited mirror writing. When she was 10 years old we found, through a fortunate coincidence and sheer perseverance, that the images in her eyes did not coincide and that she was always seeing two (unrelated) words halfway superposed. She began wearing strong prismatic glasses that made the images coincide, and from then on her reading improved dramatically.

Here we have a case of dyslexia where poor visual processing *was* involved and the cure was not "consistent remedial practice in reading." This letter is intended not to detract from Vellutino's findings but to warn against oversimplifying, against the misplaced feeling that we now have dyslexia licked and against the automatic application of his recommended remedies to all dyslexics.

DICK GRUNE

ARWEN GRUNE

Amstelveen, The Netherlands

To the Editors:

Frank R. Vellutino presents a one-dimensional "cure" for dyslexia, a problem so complex and intricate that defining it scientifically has become an exercise in futility. Even Vellutino refers to it as "a generic term that has come to refer to an extraordinary difficulty experienced by otherwise normal children in learning to identify printed words, presumably as a result of constitutional deficiencies." Based on this very general description and a series of implicative and suggestive phrases such as "appears to be," "based on the assumption," "believed to take place," Vellutino disposes of every educational construct relating to dyslexia, giving two or three lines to each one. He then argues, "We believe the problem lies in the linguistic domain."

Vellutino completely ignores the developmental changes in dyslexic children. Research I have done alone and with R. Mozlin and D. A. Rumpf revealed that the expected relations of sensory-motor, intersensory and spatial skills to reading in the primary and intermediate elementary school grades are indeed quite varied—a point that Vellutino assiduously avoids. Different skills are dominant in their influence on learning at different ages. As children move into the middle elementary grades, ceiling effects tend to take over, and the value of distinguishing between readers and nonreaders, as Vellutino does, is diminished significantly. Spatial skills, which have the dominant effect on learning to read in kindergarten and first grade, become coequal with verbal and linguistic skills in their influence on reading four years later.

J. A. Fletcher and Paul Satz of the University of California at Los Angeles point out that in Vellutino's studies the WISC (Wechsler Intelligence Scales for Children) Verbal I.Q. scale, an excellent measure of linguistic ability, revealed differences between older reading groups but not younger ones. The second-grade disabled readers, on the other hand, displayed a depressed WISC Spatial Performance I.Q. with respect to Verbal I.Q., suggesting visual-spatial deficits.

According to Vellutino, "More conventional approaches to remedial instruction have greater success, particularly in educational settings equipped to provide dyslexics with the type and amount of help they need." If this were indeed the case, poor readers would not make up close to 15 percent of the school population, or several million children.

HAROLD A. SOLAN

State College of Optometry
New York City

To the Editors:

Frank R. Vellutino's contention that dyslexia is a complex linguistic deficiency and not, as is widely believed, a defect in visual perception prompts me to report observations consistent with his approach.

Because German is my mother tongue, I have sometimes been irritated by the frequency of misspelled German words in English publications and in particular in the scientific literature. Frequent among these mistakes are a couple of inversions that in someone writing in his native language might be interpreted as symptoms of dyslexia, namely the confusion of *ei* and *ie*. An extreme example is a page of references, in a journal published by one of the established university presses, that contained among several other mistakes no fewer than eight such inversions. The same mistakes also occur in the daily press; *Wien* is spelled *Wein* and vice versa, *Beine* is *Biene, Friede* is *Freide, Zeit* is *Ziet* and so on.

To a person copying from a foreign text, an unfamiliar word may appear either as a meaningless sequence of letters, which he must copy piecemeal, or it may remind him of some word in his own language, in which case he will try to make the transfer in larger chunks or as a whole. The *e-i* inversions probably arise in the second way and can (following Vellutino's argument) be explained as reflecting nonvisual components of a linguistic apparatus, in this case phonemes in the copying process.

In English *ei* is pronounced as in the words *vein* or *skein,* whereas in German it is pronounced as in the English words *mine* or *light.* The inverse sequence *ie* in English is sometimes pronounced as a long *i,* as in *flies* or *flight,* but in others it represents a long *e,* as in *chief* or *brief;* in German *ie* is always pronounced as in *chief.* Adding to the confusion are words such as *vine* and *Wein,* which are pronounced the same way and mean the same thing but are spelled differently, words that are spelled the same way but that are pronounced differently and mean different things, such as *kind* and *Kind* (German for *child*), and other permutations. Thus it is not surprising that an English speaker having at best a rudimentary "internal" dictionary of German gets lost in the process of copying. A similar defect in the linguistic reference system may well be one cause of dyslexia. I wonder how many misprints this letter will contain?

The mimicking of a largely hereditary deficiency, such as dyslexia, as a result of special environmental circumstances is also observed in music, a nonverbal system of communication. With a colleague I have found that people who are unfamiliar with Western popular music do not spot "wrong notes" in it and make mistakes similar to those made by the congenitally tone deaf.

H. KALMUS

University College London

(*Vellutino replies on page 10.*)

© 1987 SCIENTIFIC AMERICAN, INC

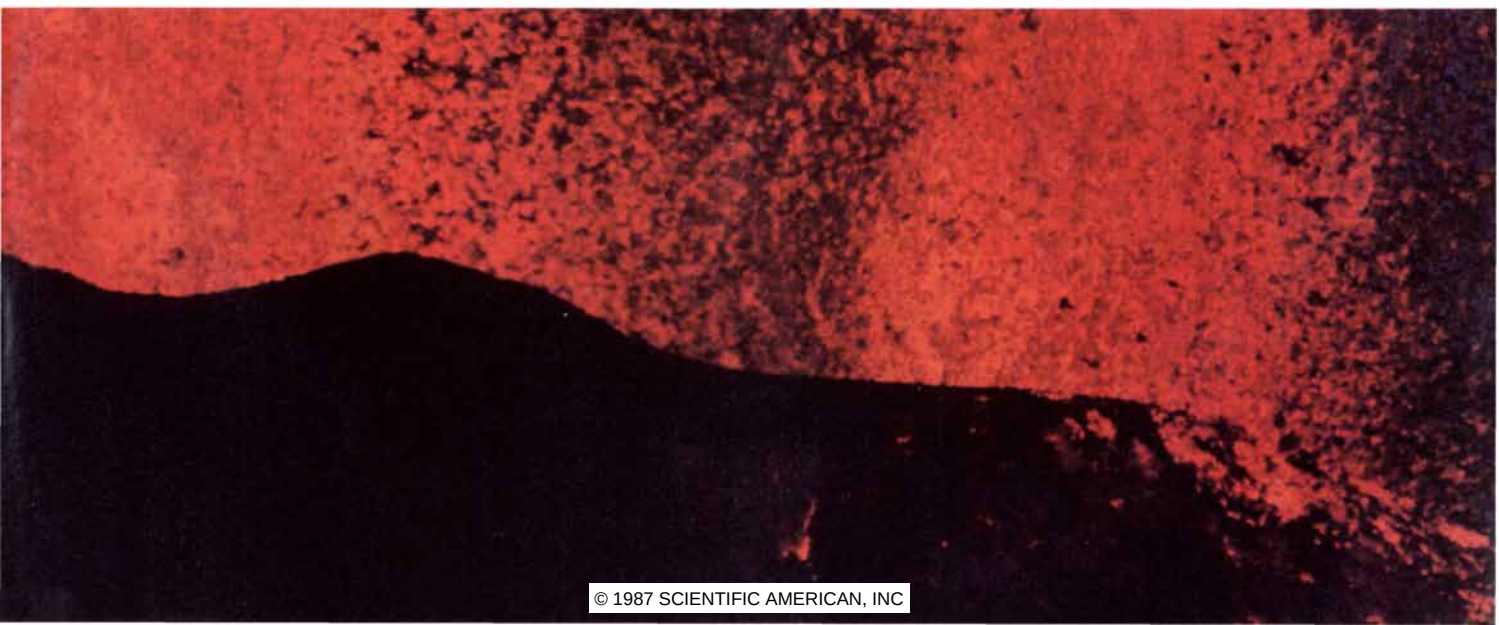
a one-centimeter movement of the earth's crust can sometimes be illuminating.

Allied Signal
Technologies

technology: Bendix laser ranging, for NASA's laser tracking network. Beamed from earth to satellite, lasers measure earth's crust movements as little as one centimeter, over a 3,000 mile range. From Allied-Signal.

benefit: Helps scientists gather data to find a way to predict natural disasters such as earthquakes, volcanic eruptions, and tidal waves.

future: An increasing number of countries around the world will use Bendix laser ranging. Allied-Signal advanced technologies—focused on aerospace, automotive products, and engineered materials. For investor brochure, call 1-800-243-8160.



© 1987 SCIENTIFIC AMERICAN, INC

To the Editors:

Dr. Richard Grune and Arwen Grune misinterpreted the connection between mirror writing and reading disability. The article did not indicate that mirror writing is a symptom of dyslexia but rather that mirror writing is not intrinsically related to reading difficulties of any type or origin except in clinical lore. Reversals of letters and words are *normally* observed in young children first learning to write because they become aware of the need for positional and directional constancy only as they learn. That is, they acquire motor "programs" (to use a computer metaphor) for producing letters and words before they learn the importance of executing these programs in a way that maintains correct orientations and sequences. For example, a child may learn to print the letter *N* using a continuous zigzag program. If he or she initiates the program going from left to right and from bottom to top, the *N* will be oriented correctly, but initiating the program either from right to left or from top to bottom results in an incorrectly oriented letter. Provided young children learn to read and spell successfully, they have many opportunities to practice writing letters and words in the correct orientations and sequences. But if they have difficulty learning to read and spell, they have few such opportunities. Mirror writing may persist in these children, but not because of inherent spatial deficits.

With respect to Arwen Grune's experience, it will suffice to point out that any optical deficiency should of course be corrected. What is at issue, however, is the contention of some vision specialists that reading problems may be caused by presumed deficiencies in ocular functioning that lead to "visual perceptual problems" even in poor readers with good visual acuity. Such problems, it is argued, are manifested in symptom patterns that include reversals of letters and words, directional-scanning and sequencing problems and visual-motor impairments in writing. The remedy for such deficiencies is said to be visual training to facilitate binocular coordination, visual tracking, ocular motility and so on.

Three points can be made to refute these claims. First, carefully controlled laboratory studies (including several of our own) have consistently shown that anomalous patterns such as reversal errors and directional-scanning problems are the consequences of reading difficulties rather than benchmarks of perceptual problems that cause reading difficulties. Second, there is now abundant evidence that the variety and number of ocular defects in young children is no greater in poor readers than in normal readers. This means, in effect, that many such deficiencies can be "tolerated" with little effect on the ability to learn to read. Finally, the balance of evidence from the very few controlled studies of the effects of visual training on reading ability weighs heavily against such training as a remedy for reading problems.

Dr. Solan disagrees, asserting that visual-motor and other somatosensory abilities are the primary determinants of success in beginning reading, whereas language ability is more important for success at later stages of skills acquisition. To support this position (initially articulated by Paul Satz) Solan makes specific reference to two studies he conducted, which presumably address the question.

The study done with Mozlin and Rumpf found moderately high correlations between performance on a "school readiness" test administered to kindergartners (ages five and six) and performance on tests of remembering briefly exposed numbers, matching word forms, placing small pegs in slots and matching rhythmic patterns with corresponding dot patterns. The school readiness test, however, tested similar skills (perceptual speed, facility with numbers, spatial skill, verbal comprehension and auditory discrimination), and there were no measures of rudimentary reading skills such as letter and word identification. In addition the subjects' I.Q.'s varied between 72 and 145, and the study did not control for intelligence. Hence the correlations could have reflected the similarities in the tasks and/or the differences in intelligence.

In the other study Solan refers to, randomly selected fourth and fifth graders were given all the verbal subtests and one spatial subtest from the WISC-R (the revised WISC), along with other measures of visual-spatial and visual-motor ability. They were also given measures of word identification and reading comprehension. Because there were moderately high correlations between the reading and the visual-spatial measures as well as between the reading and the verbal measures, Solan concludes that language ability and visual-spatial ability are both important determinants of skill in reading in children in the intermediate age and grade levels.

The subjects for this study were selected at random, however, and the sample probably included few poor readers. Hence the study did not actually compare poor and normal readers. Moreover, since most of the tasks were taken from the WISC-R, they cannot be called pure measures of either language or visual-spatial ability; such tasks are complex and often require higher level cognitive skills including judgment, reasoning and problem solving. Solan's data may therefore reflect little more than the reasonably high correlations typically observed between general intelligence and reading achievement.

Solan refers to commentary by Fletcher and Satz about intelligence-test results in some of my own studies. Like Solan, these investigators have made the mistake of equating the verbal and nonverbal parts of the intelligence test we used (the WISC-R) with language and spatial ability respectively. This is simply not legitimate from the standpoint of psychometrics. I have debated these issues with Fletcher and Satz elsewhere (Vellutino, *Journal of Learning Disabilities*, Vol. 12, No. 3, pages 160–167; 1979).

Finally, Dr. Kalmus' comments and examples illustrate quite nicely the point that the orientation and sequencing errors often observed in the reading and spelling of poor readers probably result from deficiencies in the use of linguistic knowledge to analyze the writing system rather than from any basic (neurological) deficiencies in visually perceiving and sequencing letters. It is the linguistic components (semantic, phonological and syntactic) of printed words, not their visual components, that imbue them with meaning and determine how one should group and order their constituent letters. If a child has difficulty acquiring linguistic knowledge, he or she will have difficulty learning how printed letters and words "map" onto language and will therefore be subject to the kinds of orientation and sequencing errors that have long masqueraded as symptoms of a perceptual deficiency. Kalmus' letter shows how existing sound-print conventions can lead to superficial processing of the type that results in "misperception" and spelling errors. The dyslexic's problem is that the difficulties in acquiring sound-print conventions result in the same types of processing and hence the same types of errors.

FRANK R. VELLUTINO

© 1987 SCIENTIFIC AMERICAN, INC

A composite material made of graphite epoxy is ideal for use in space because it is stronger than steel yet lighter than aluminum. Unlike metal, it does not expand from exposure to heat or cold, and its reduced weight translates into lower costs and larger payloads. For example, saving just one pound in a space vehicle means a reduced launch cost of several thousand dollars or added fuel for longer operations. Until now, however, graphite composites have proved difficult to mold to shapes more complex than a simple cylinder. But research and development by Hughes Aircraft Company has opened the way for the fabrication of a variety of new forms, including tubes with integrated end fittings in a one-piece design, support beams, and ring structures up to seven feet in diameter.

The HR-3000, a new generation version of the Hughes Air Defense Radar (HADR), can detect and precisely locate fighter-sized aircraft more than 470 kilometers away. The phased array radar is designed for air traffic control in peacetime and can automatically detect wartime threats that appear simultaneously from several directions, even in the presence of heavy environmental clutter and severe electronic interference. The HR-3000 system is in full production at Hughes and will be used in Portugal and integrated with Hughes-built NATO Air Defense Ground Environment (NADGE) in Italy, Greece, and Turkey.

Upgraded flight simulators will depict mission imagery realistically for U.S. Navy F/A-18 pilots. Hughes is modifying its Weapons Tactics Trainers (WTT) to project high-resolution, full color, real-time simulated images of terrain features and man-made structures on 360-degree field-of-view dome screens that surround the trainer's cockpit. The out-the-window scenes will be produced by advanced computer imaging technology, using a digital database that represents 70,000 square miles of western Arizona and southern California terrain. The new simulated visual system will allow pilots to safely practice a host of complex combat scenarios without expending fuel or weapons.

An aircraft collision avoidance function is part of the new air traffic control system in the Republic of Korea. One of the world's most sophisticated civilian air traffic control (ATC) systems warns controllers when aircraft fly in converging paths, descend below a safe altitude, or approach restricted airspace. Radar data and flight processing functions are automated and combined in the system, enabling controllers to move traffic safely and expeditiously. The system monitors data simultaneously from multiple overlapping radar networks, creating a tracking picture significantly more reliable than current systems. When a potential conflict is identified, the system automatically provides an early visual and audio alert on the air traffic controller's console. Under development for three years, the ATC system and the aircraft collision avoidance function were both developed and built by Hughes. The system provides complete control of South Korean airspace, including en route and airport approach control throughout the nation and over ocean areas.

An ultramodern facility spanning 1.75 million square feet is the showcase where outstanding Hughes engineering combines advanced manufacturing techniques and production processes. Our complex is complete, so we're looking for experienced and graduating engineers to work on such programs as infrared thermal imaging systems, laser rangefinders and designators, artificial intelligence, signal processing, analog, VLSI, design, component qualification, hybrid microcircuitry, and focal planes. Send your resume to Hughes Electro-Optical and Data Systems Group, Professional Employment, P.O. Box 913, E8/G101, Dept. S2, El Segundo, CA 90245. EOE. U.S. citizenship required.

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

© 1987 Hughes Aircraft Company

HUGHES
AIRCRAFT COMPANY

Subsidiary of GM Hughes Electronics

© 1987 SCIENTIFIC AMERICAN, INC

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

SEPTEMBER, 1937: "Mars continues to be a desert, defying astronomers' biggest telescopes and most delicate instruments to find any trace of water vapor on the rust red surface of its middle part. So report Walter S. Adams and Theodore Dunham, Jr., of the Mount Wilson Observatory. Last April, Mars was in an especially favorable position for observation. The astronomers turned the great 100-inch telescope on the planet. Dark absorption lines appeared in the spectrum, part of them due to the 'soaking up' of the planet's light by water vapor in the atmosphere of the earth. Particularly critical study was made to see if any absorption took place in water vapor in the atmosphere of Mars. Of this the astronomers reported they could find 'no evidence whatever.'"

"In physics class it was taught that radio waves travel in an ever expanding circle, like the ripples created when a pebble is tossed into still water. They do, when left to their own devices, but not being satisfied with anything as simple as that, engineers have devised ways to make the waves radiate in the form of a fan, a shamrock, a four-leaf clover, a spatula, a double watermelon or an airplane propeller. These are only a few examples of what are called field patterns that are determined by the design of the transmitter antennas. Broadcasting has ceased to be broadcast and is now a scientific distribution of radio energy."

"Work on the All-American Canal is going forward so rapidly that the great 80-mile ditch and its appurtenant features will be ready to deliver irrigation water to the Imperial Valley in southeastern California sometime during the first half of 1938. The three principal features of the project are the diversion dam across the Colorado by which the river water will be impounded and directed into the headworks on the California side of the river; the immense desilting plant that will remove 70 percent of the solid matter in suspension in the water, and the canal itself."

"The racing yacht *Ranger,* which will defend the America's Cup, embodies the latest principles of yacht design and construction. *Ranger* was designed by W. Starling Burgess, designer of the last America's Cup winner, *Rainbow,* as well as the previous *Enterprise.* Mr. Burgess' father before him designed three America's Cup winners."

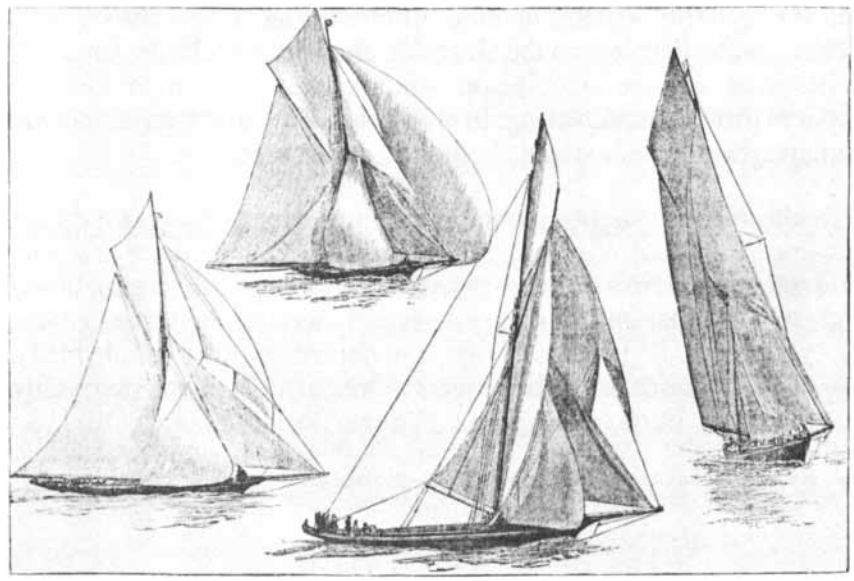
"What is reputed to be the first drive-in bank in the world has recently been opened at Vernon, a suburb of Los Angeles. The motorist merely drives into the bank building and transacts his business without even moving from his seat."

SCIENTIFIC AMERICAN

SEPTEMBER, 1887: "Few, if any, branches of scientific inquiry have done so much for the 'greatest good of the greatest number' as those which have unfolded and developed the principles of hygiene and sanitation, and taught us that the preservation of good health and the prolongation of life both chiefly depend on a supply of pure air and pure water. It has over and over again been proved that in such crowded cities as our own the most malignant forms of infection are chiefly propagated by spores or germs floating in the air or carried into our drinking water. Although a great deal of attention has been devoted to the analysis of both these elements, there is still considerable room for the simplification of processes and the popularization of easy and reliable methods of investigation."

"The absence of a parcel post service, both inland and foreign, is a disgrace to the government of the United States. The U.S. Postmaster-General seems, from his annual report, to have made a small effort to give his fellow citizens such use of the mails as is enjoyed by people abroad, but, as he naively says, the express companies opposed the idea so strongly that he was obliged to abandon it. It is not unnatural that the express companies should oppose a plan for transporting packages by the mails at a small fraction of the rates they demand for the same service, but there is something to be considered in the administration of a public office besides the interests of those companies. Their opposition ought not to be regarded for a moment in comparison with the great benefits of parcel post."

"The yacht *Thistle,* constructed in Scotland with a view to competing for the America's Cup, arrived in New York on the 16th of August. The vessel attracts much attention. Opinion is about evenly divided on whether she will be able to beat the *Volunteer,* the new yacht with which she will compete for the great prize in October. In nearly all of the several races the *Thistle* has sailed in England she has been the winner."



***Views of the yacht* Thistle**

© 1987 SCIENTIFIC AMERICAN, INC

© 1987 SCIENTIFIC AMERICAN, INC

Hitachi's wide-ranging automotive technologies include car audio, the Satellite Drive Information System featured on Nissan's CUE-X concept car, a car telephone, and a microcomputer engine control system.

© 1987 SCIENTIFIC AMERICAN, INC

AUTOMOTIVE

Automobiles should be more than safe, comfortable machines. They should also be able to communicate with the world around them.

Recent advances in car electronics technology have been remarkable. They've not only improved basic functions such as engine control, they're now being seen in man-machine interfaces providing more comfort and operating ease, and even in communications with the surrounding world. Down the road there are things even more exciting.

Hitachi's scientists and engineers are at work on a Multi Information System using a color thin filter transistor LCD to display operating information, road maps and a navigational system using these maps. With this system a driver could obtain a variety of driving information simply by touching the display screen. Eventually, he'll be able to issue verbal commands to, for instance, regulate the temperature within his car. Hitachi electronics and semiconductor technology can also bring free communication with the outside and determine a car's exact location through use of Global Positioning System satellites.

Hitachi has also developed a highly acclaimed hot wire air flow sensor used in engine management. It helps achieve the diametrically opposed goals of maximum power and fuel economy. And we've created many other superior products for driving control, suspension control, air conditioning and audio.

We link technology to human needs; and believe that our special knowledge will create new, highly sophisticated functions that are also easy to operate. Our goal in automotive electronics — and medicine, energy and consumer electronics as well — is to create and put into practice innovations that will improve the quality of life the world around.

HITACHI

Hitachi, Ltd. Tokyo, Japan

© 1987 SCIENTIFIC AMERICAN, INC

THE AUTHORS

C. KUMAR N. PATEL and NICOLAAS BLOEMBERGEN ("Strategic Defense and Directed-Energy Weapons") cochaired the panel created to evaluate the feasibility of directed-energy weapons for the American Physical Society. Patel, who heads the physics and academic affairs division of the AT&T Bell Laboratories, was born in India and received his undergraduate degree at the University of Poona in 1958. He holds an M.S. (1959) and a Ph.D. (1961) from Stanford University. Soon after joining Bell Laboratories in 1961, Patel invented the carbon dioxide laser. Bloembergen, who is Gerhard Gade University Professor at Harvard University, was one of the recipients of the 1981 Nobel prize in physics. He was born in the Netherlands, obtained his D.Phil. in 1948 from the State University of Leiden and became an associate professor at Harvard in 1951. Bloembergen won a National Medal of Science in 1974.

ALAN DRESSLER ("The Large-Scale Streaming of Galaxies") is an astronomer at the Carnegie Institution of Washington's Mount Wilson and Las Campanas Observatories. He has a bachelor's degree from the University of California at Berkeley and in 1976 was awarded a Ph.D. in astronomy and astrophysics by the University of California at Santa Cruz. After five years as a postdoctoral fellow at what were then known as the Hale Observatories he took up a permanent position. Dressler specializes in optical observations of other galaxies and is particularly interested in galaxy formation and evolution.

HAROLD VARMUS ("Reverse Transcription") became interested in RNA tumor viruses (now known as retroviruses) while working as a clinical associate at the National Institutes of Health in the late 1960's. Currently American Cancer Society Professor of Molecular Virology at the University of California at San Francisco, he began his academic career studying English literature, in which he earned a B.A. at Amherst College in 1961 and an M.A. at Harvard University in 1962. He enrolled in medical school at Columbia University in 1962 and on receiving his M.D. in 1966 began an internship at Columbia-Presbyterian Hospital in New York. Varmus went to the National Institutes of Health in 1968 before moving to U.C.S.F. in 1970.

JAMES L. DYE ("Electrides") chairs the department of chemistry at Michigan State University. He joined the faculty after getting an A.B. in 1949 at Gustavus Adolphus College and a Ph.D. in 1953 from Iowa State University. In 1975 and 1976 he worked as a Guggenheim Fellow and a Fulbright Research Scientist at the University of Strasbourg, and he also spent part of 1982 and 1983 on sabbatical at Bell Laboratories. Dye has directed Michigan State's Center for Fundamental Materials Research and is a member of the American Chemical Society.

SPENCER C. H. BARRETT ("Mimicry in Plants") is associate professor of botany at the University of Toronto. His specialties—the evolution of mating systems in plants and the genetics of weeds—grew out of an expedition to Swaziland required for his undergraduate degree in agricultural botany. After receiving his B.Sc. in 1971 at the University of Reading, he spent a year in Brazil collecting seeds for the genetic-resources conservation program of the Royal Society of London. He then went to the University of California at Berkeley, where he was awarded his Ph.D. in 1977. Barrett has been at Toronto since receiving his doctoral degree.

MARK HORTON ("The Swahili Corridor") has directed numerous expeditions to Shanga and Zanzibar since 1980. He holds a B.A. (1978) and Ph.D. (1984) from the University of Cambridge and now works as a research fellow in archaeology at the University of Oxford. Along with his excavations in Africa, which are sponsored by the British Institute in Eastern Africa, Horton has explored sites in Sri Lanka and Central America, and his enthusiasm for industrial archaeology frequently leads him to dig in British soil as well.

GEORGE A. MILLER and PATRICIA M. GILDEA ("How Children Learn Words") began their collaboration when Gildea was doing postgraduate work at Princeton University. Miller is currently James S. McDonnell Distinguished University Professor of Psychology at Princeton, where he has been on the faculty since 1979. After getting his bachelor's and master's degrees at the University of Alabama he earned a Ph.D. from Harvard University in 1946. He worked as a research fellow in Harvard's Psycho-Acoustic Laboratory until 1948, when he was made assistant professor of psychology. Between 1951 and 1955 he taught at the Massachusetts Institute of Technology; in 1955 he returned to Harvard, where, in 1960, he and Jerome S. Bruner founded the Center for Cognitive Study. Miller went to Rockefeller University in 1968 and to Princeton 11 years later. Gildea worked on the "dictionary project" through a postdoctoral position she held at Bell Communications Research in 1984. Currently assistant professor of psychology at Rutgers University, she has a B.A. (1979) from Sarah Lawrence College and an M.A. (1981) from Princeton. Her interest in vocabulary development grew out of research for her Ph.D., which she received from Princeton in 1984. Gildea went to Rutgers in 1985.

RICHARD E. BALZHISER and KURT E. YEAGER ("Coal-fired Power Plants for the Future") have been at the Electric Power Research Institute (EPRI) in Palo Alto since the early 1970's. Balzhiser, who went to the institute in 1973, heads its research and planning activities as executive vice-president. His bachelor's (1955), master's (1956) and Ph.D. (1961) are from the University of Michigan, where he was made professor of chemical engineering in 1961. He taught there until 1971, when he joined the White House Office of Science and Technology as assistant director. Balzhiser moved to EPRI two years later. Yeager, who is vice-president in charge of EPRI's coal-combustion-systems division, went to the institute in 1974. He got a B.S. in 1961 at Kenyon College and an M.S. in 1964 at the University of California at Davis. While serving in the U.S. Air Force he studied fluid dynamics and mechanical engineering at the Technical Application Center. In 1968 he joined the MITRE Corporation, where he later held a position in the environmental-systems department. Yeager was appointed director of energy research and development planning for the Environmental Protection Agency's Office of Research in 1972 and then moved to EPRI. Both authors would like to thank the staff members of EPRI who helped in the preparation of their article.

© 1987 SCIENTIFIC AMERICAN, INC

Nuclear energy helped America achieve its energy balance. Is it a balance we can keep?

The 1973 Arab oil embargo forced America to turn to alternatives to foreign oil. Reliable alternatives. America increased its use of electricity from nuclear energy and coal and began to make important strides toward energy independence.

We have since let our guard down. Oil imports are rising steadily and now rival 1973's. The implications of this foreign dependence are clear. So are the solutions.

A dangerous rise in oil imports

America imported four million barrels of oil a day in 1985. In 1986, that figure jumped to over five million barrels a day. By 1990, we will most likely rely on imports for nearly half our needs. Some say as much as 75%. Compare that to 35% in 1973.

What happens when we become too dependent on foreign sources? We lose our balance. It's the first misstep toward losing our energy security. In 1973, that meant short supplies, long gas lines, expensive fuels and critical damage to our economy.

A reliable supply of nuclear electricity

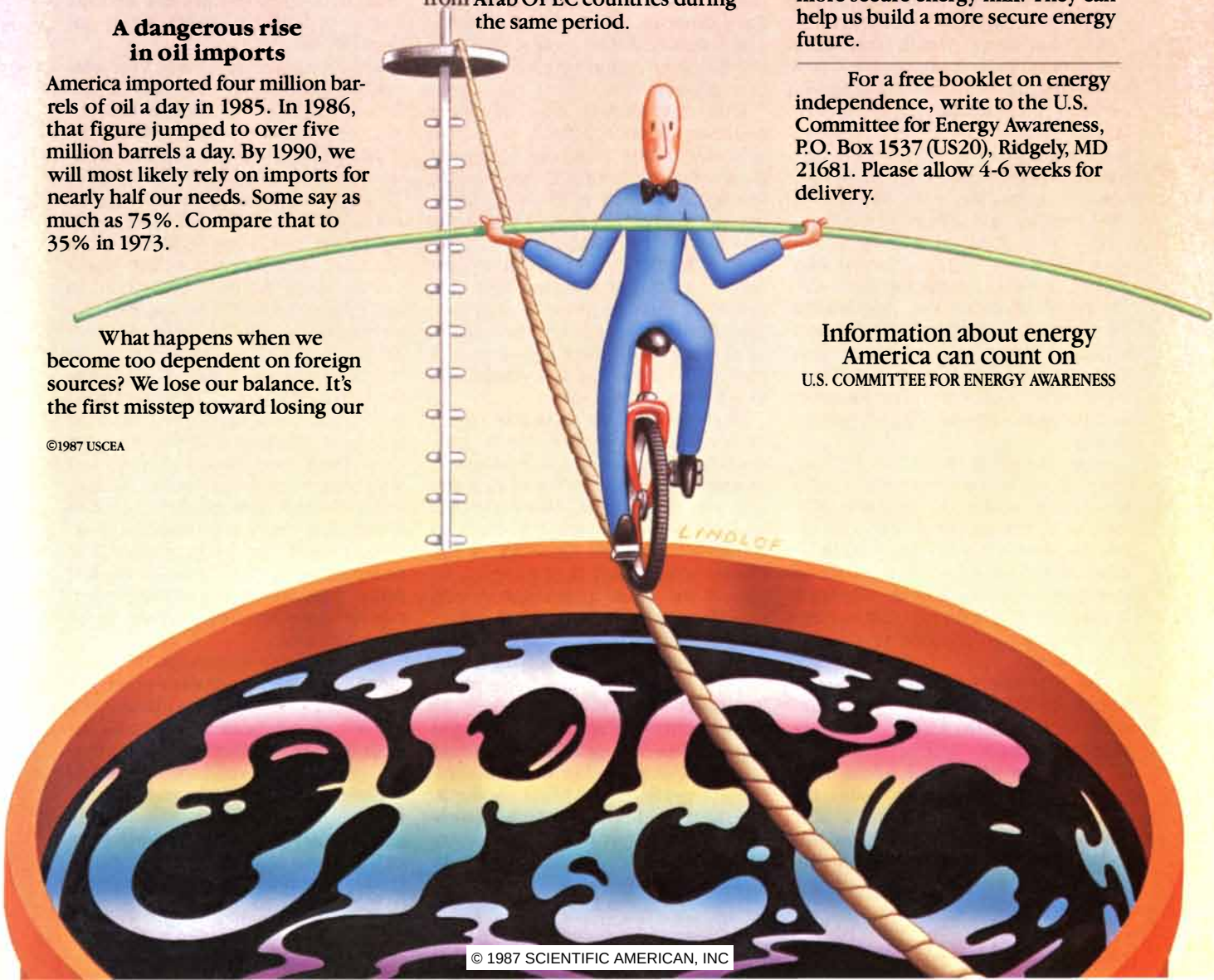
America has over one-fourth of the world's uranium. We have over 100 plants to convert it to electricity. According to energy analysts at Science Concepts, Inc., U.S. nuclear plants saved over two billion barrels of oil between 1973 and 1986. That's roughly one-fourth of the total amount of oil imported from Arab OPEC countries during the same period.

And, while our use of oil and natural gas is down from 1973 levels, we now use about 45% more coal and almost 400% more nuclear energy than we did then.

Nuclear energy for a secure future

Obviously, nuclear energy can't completely replace oil here. And our own limited oil resources will force us to continue to rely on foreign suppliers. The good news is nuclear energy and coal, America's two leading sources of electricity, have helped us establish a more secure energy mix. They can help us build a more secure energy future.

For a free booklet on energy independence, write to the U.S. Committee for Energy Awareness, P.O. Box 1537 (US20), Ridgely, MD 21681. Please allow 4-6 weeks for delivery.



Information about energy America can count on

U.S. COMMITTEE FOR ENERGY AWARENESS

©1987 USCEA

© 1987 SCIENTIFIC AMERICAN, INC

SCIENCE AND THE CITIZEN

Two Steps Forward...

Many scientists perceive a lack of direction in the space program

Since the National Aeronautics and Space Administration's solar-system exploration committee was created in 1980, "for every two steps forward we have slipped back three," David Morrison, the committee's chairman, wrote in a recent letter to NASA administrator James C. Fletcher. While plans for manned missions are in limbo (the Reagan Administration has not responded to last year's recommendation from the National Commission on Space that manned exploration of Mars be adopted as a national goal, and the plan to build a space station remains contentious) unmanned space science is faring even worse.

Morrison's letter, which constitutes his committee's final report, complains that most of the committee's earlier recommendations for a revitalized planetary-science program have been ignored. He points out that NASA's budget request for planetary science, a modest $310 million in 1988, is less than a fourth of the annual expenditure for planetary science in the early 1970's. The reduced access to space stemming from the *Challenger* disaster has aggravated the "lack of commitment." Morrison estimates that a 100 percent budget increase over three years is needed, along with a "specific commitment and well-understood long-range objectives."

After launching the planetary missions already approved—the *Galileo* mission to Jupiter, the *Magellan* mission to Venus and the *Mars Observer*—the agency's next order of business, Morrison wrote, should be to start construction of the *Mariner Mark II* spacecraft, a general-purpose device that can be adapted to a range of missions.

The first *Mariner Mark II* mission will probably be the Comet Rendezvous and Asteroid Flyby, which its supporters hope will be launched in 1994. The spacecraft would first approach an asteroid and later fly alongside the comet Tempel 2 and fire a penetrator into it. If work on the *Mariner Mark II* does not begin this year, Morrison says, the craft will not be ready in time, and there will be no other opportunity for a comet rendezvous in this century. As the second mission for the new Mariner spacecraft, Morrison recommends Cassini: a voyage to Saturn and its moon Titan, to be undertaken with the European Space Agency. Morrison warns that delay will jeopardize the Cassini mission by chilling European interest.

William L. Quaide, chief of NASA's planetary-science branch, concedes Morrison's last point but defends NASA's record, pointing out that the agency has been busy reengineering the *Galileo* mission. That probe was to have been launched from the space shuttle by a Centaur upper-stage rocket, but the plan was canceled last year on grounds of safety. The Inertial Upper Stage that will be used in place of the Centaur is less powerful, requiring a complicated gravity-assist trajectory.

At the same time NASA has begun to investigate possible future missions for the manned space program, establishing a new Office of Exploration for the purpose. Many officials see manned exploration of Mars as a long-term target. Some think the closed life-support technologies required for such a mission—which would take months in each direction—should first be tested in the relative proximity of the moon. A permanent lunar base would also make it possible to carry out studies of the physiological effects of prolonged exposure to low gravity and would provide a valuable site for astronomical observations and other research. The idea seems to be gaining momentum: NASA has allocated $1.2 million this year for studies of how to build a lunar base. —*Tim Beardsley*

Chemical Warfare

A truce between environmentalists and pesticide makers unravels

A fight is brewing as Congress tries once again to rewrite the principal law controlling the licensing of pesticides: the Federal Insecticide, Fungicide, and Rodenticide Act, or FIFRA. New far-reaching controls on pesticides that find their way into groundwater have been proposed in the Senate and are likely to be opposed by the National Agricultural Chemicals Association (NACA). The groundwater controls go beyond those agreed to last year by an alliance of the NACA and the Campaign for Pesticide Reform, an umbrella group of environmental organizations and consumer groups, and the short-lived truce seems likely to break down.

The compromise of last year was reached after the environmental lobby agreed to support the manufacturers' long-standing request for an extension of a pesticide patent's duration from 17 years to a maximum of 22 years, an increase meant to compensate for the time it takes a company to get marketing approval from the Environmental Protection Agency. In return the industry agreed to pay fees to defray the cost of the EPA's slow-moving effort to complete modern safety reviews of all 600 pesticide ingredients in use today. The EPA has been working at this "reregistration" since 1972 but by the end of this year will have completed the process for only three ingredients. Under last year's agreement the EPA would have been compelled to finish the task in about nine years, with industry paying $150,000 for each review. The compromise legislation fell at the last fence, however, when Congress ran out of time for considering some amendments.

This year the Campaign for Pesticide Reform is capitalizing on the election of a more sympathetic Congress by seeking to strengthen controls further. Besides the groundwater provisions (which would allow the EPA to make manufacturers themselves monitor contamination of groundwater and also to halt the use of offending pesticides) the new proposals would increase the reregistra-

© 1987 SCIENTIFIC AMERICAN, INC

BREAKTHROUGH: BAKE COMPOSITES AND NEVER BOTCH A BATCH.

Strong, lightweight carbon composites that resist heat are essential for high performance jet aircraft.

Producing consistent, high-quality parts with the material, however, has been a delicate and somewhat unpredictable process. Layers of strong, light composite fabric must be slowly and skillfully baked and cured into precise aerodynamic shapes.

To make sure that every batch turns out right, we've developed a fiber-optic sensor probe to measure quality right in the curing ovens so that our technicians know exactly how fast the cure is progressing and when each batch of parts is perfectly done. Inconsistencies from batch to batch are minimized. Optimum properties are maintained. Waste is cut. Time and money are saved.

We're creating breakthroughs that make a difference in the way things work and the way people live.

We're McDonnell Douglas.

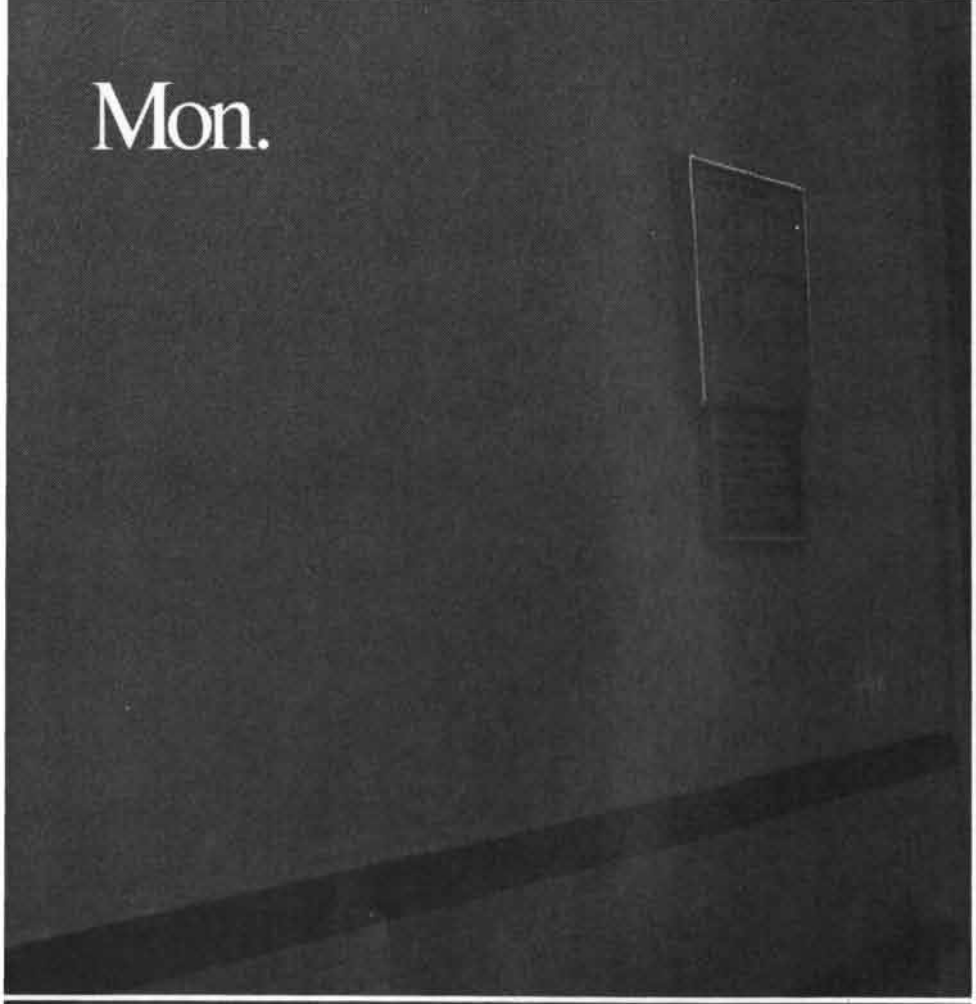
For more information, write: Advanced Materials, McDonnell Douglas, Box 14526, St. Louis, MO 63178

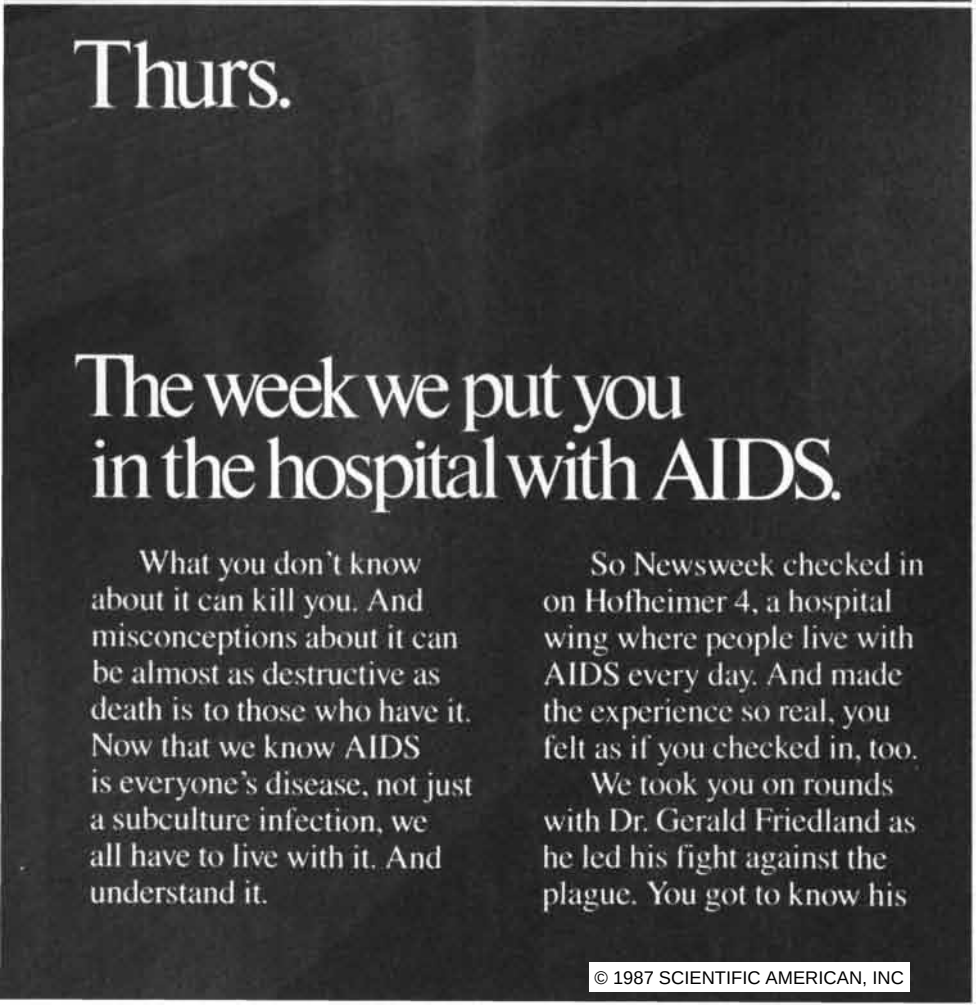
MCDONNELL DOUGLAS

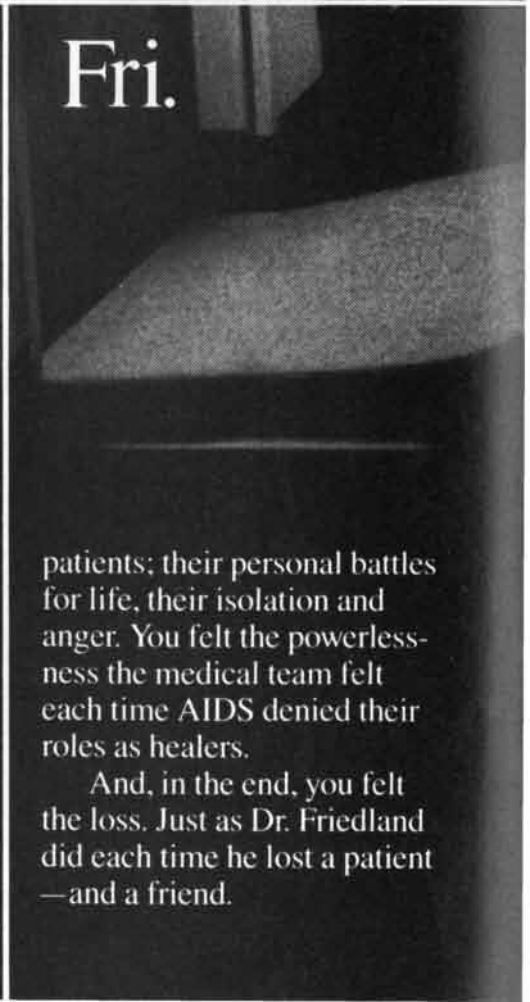
MILITARY & COMMERCIAL AIRCRAFT SPACECRAFT & MISSILES HEALTHCARE SYSTEMS INFORMATION SYSTEMS HELICOPTERS FINANCING

© 1986 McDonnell Douglas Corporation

© 1987 SCIENTIFIC AMERICAN, INC







The week we put you in the hospital with AIDS.

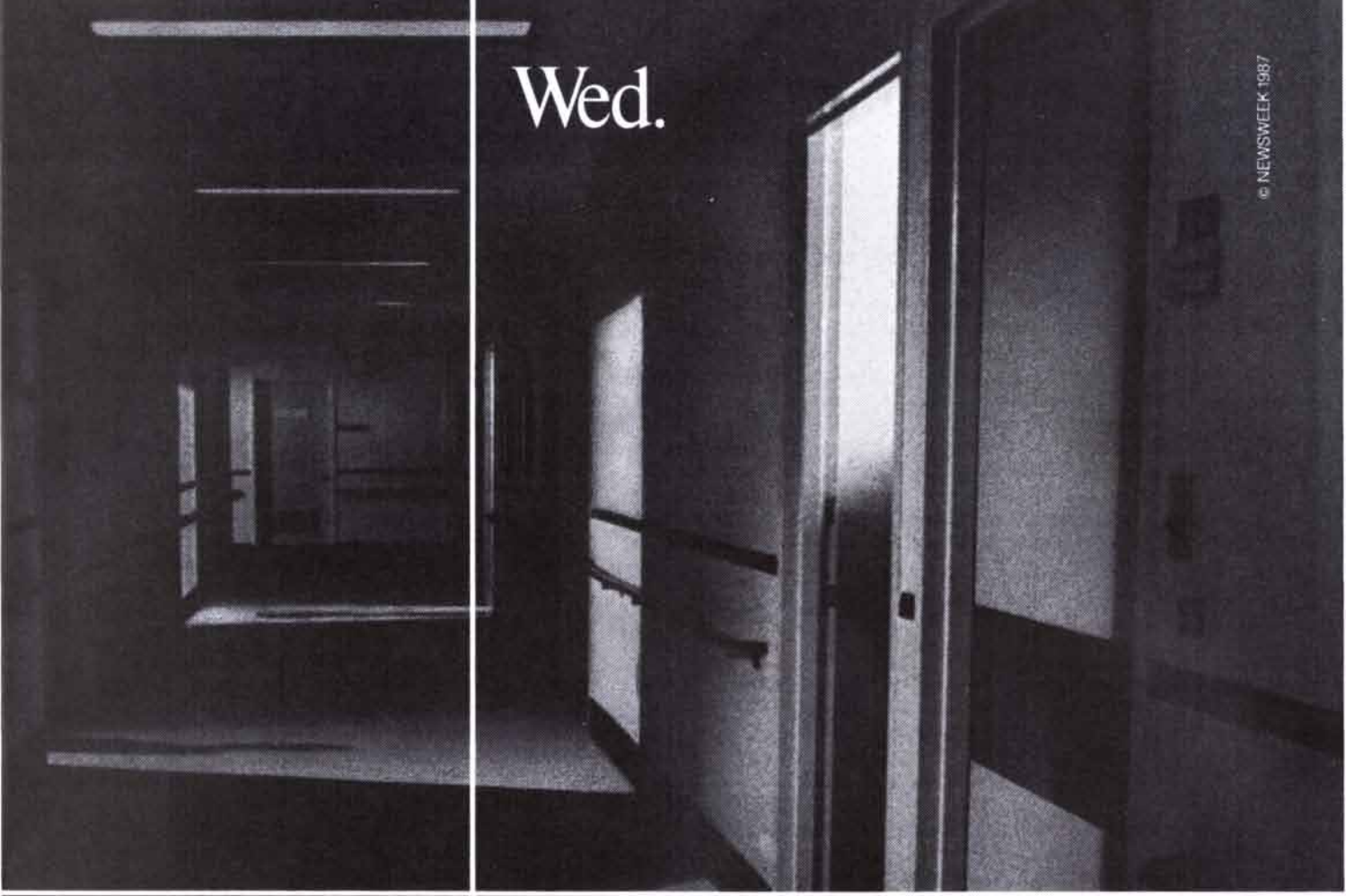
What you don't know about it can kill you. And misconceptions about it can be almost as destructive as death is to those who have it. Now that we know AIDS is everyone's disease, not just a subculture infection, we all have to live with it. And understand it.

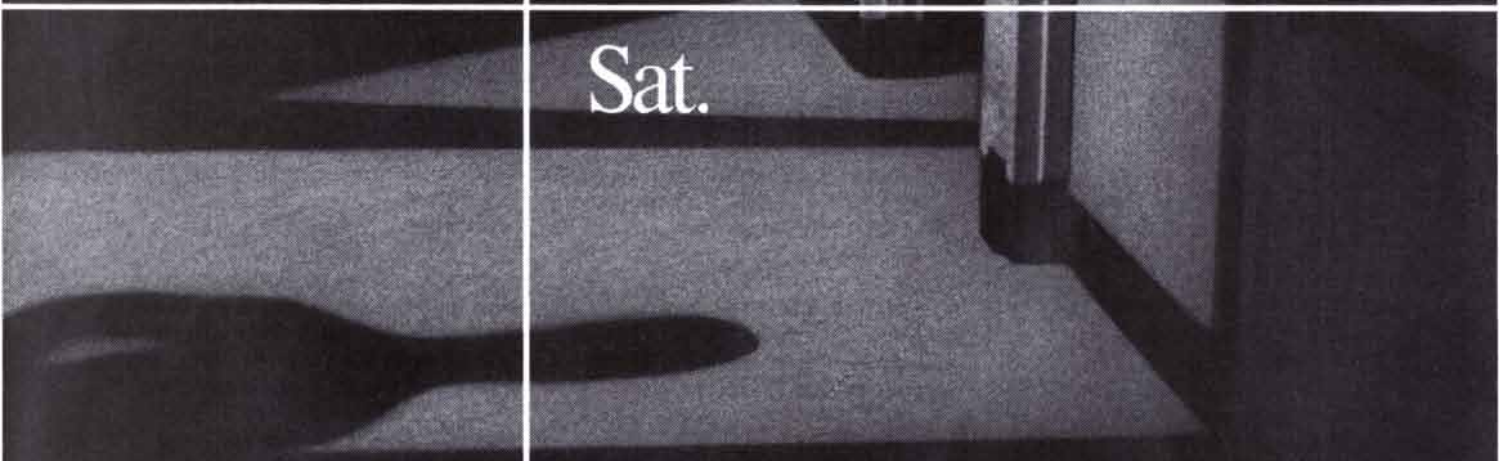
So Newsweek checked in on Hofheimer 4, a hospital wing where people live with AIDS every day. And made the experience so real, you felt as if you checked in, too.

We took you on rounds with Dr. Gerald Friedland as he led his fight against the plague. You got to know his patients; their personal battles for life, their isolation and anger. You felt the powerlessness the medical team felt each time AIDS denied their roles as healers.

And, in the end, you felt the loss. Just as Dr. Friedland did each time he lost a patient —and a friend.

© 1987 SCIENTIFIC AMERICAN, INC





Close enough to let it touch you—that's how close Newsweek gets you to all the issues that have an impact on your life.

When we reported on the chaotic, often brutal state of the Texas prison system, we went to jail. So you could see the story from the inside.

When we reported on cancer, we lived with it first. So you could experience one woman's experience with her illness as it happened.

That's why more and more people are reading Newsweek. We bring human issues home to you in a human voice. These aren't stories you just read through. These are stories you live through.

Newsweek

Why it happened. What it means.

© 1987 SCIENTIFIC AMERICAN, INC

tion fee. They would also scrap the present requirement that the EPA automatically compensate industry for the value of pesticide stocks that have to be destroyed—a "ludicrous" requirement, according to John A. Moore, the EPA's assistant administrator for pesticides and toxic substances. The NACA says it will oppose any changes that go beyond last year's compromise. This year may be the last chance for pesticide reform for some time, because congressional and presidential elections will make the passage of controversial legislation difficult.

One legal snare for the EPA that will almost surely not be abolished by any new legislation is the infamous Delaney clause, a 30-year-old statute that forbids the EPA to approve any use of a pesticide that induces tumors in laboratory animals if it is also concentrated in the course of food processing. Pesticides that are not concentrated in this way are regulated on a risk-benefit basis. The distinction may, paradoxically, increase the risk of cancer from dietary pesticide residues, according to a recent study by the National Academy of Sciences: the EPA must allow known oncogenic pesticides that do not become concentrated to be used in some applications because a less oncogenic substitute falls afoul of Delaney. Yet because of its popular reputation as a beneficent requirement, the clause is considered to be politically unmovable. *—T.M.B.*

PHYSICAL SCIENCES

Big-Bang Bashers

New research revives dissident views of the universe's formation

Few ideas in modern science have proved as robust as the theory of the big bang. The theory holds that the universe exploded into being some 15 to 20 billion years ago and that as it expanded, gravity caused relatively dense regions of matter to coalesce into stars and galaxies. The faint microwave radiation that pervades the universe, theorists say, is the afterglow of the cataclysm, and the red shifts (displacements toward the red end of the spectrum) seen in light emanating from distant objects testify to the continuing expansion of the universe. Like the theory of evolution, the big-bang model has undergone modification and refinement, but it has resisted all serious challenges.

Nevertheless, ever since the theory won general acceptance about 20 years ago a few scientists have persistently attacked some of its fundamental assumptions. One group of critics argues that electromagnetic forces generated by plasma have been more important than gravity in shaping the universe; another asserts that red shifts are not necessarily a relic of the big bang's continuing outward thrust. These groups, although long relegated to "fringe" status by mainstream astrophysicists, have been invigorated of late by new converts and new findings.

The elder statesman of the plasma dissidents is Hannes Alfvén of the Royal Institute of Technology in Stockholm. Alfvén, who won the Nobel prize for physics in 1970, believes interstellar space is filled with long filaments and other structures of plasma, that is, electrons and positively charged ions. The same electromagnetic forces that push plasmas into distinctive shapes in the laboratory, Alfvén says, caused this cosmic plasma to coalesce into galaxies, stars and planetary systems.

Alfvén believes the universe is expanding, but he speculates that the expansion is driven by the energy released when matter and antimatter meet and annihilate each other. He also believes the expansion is less dramatic than the big-bang theory proposes: a universe dominated by electromagnetic forces, he contends, could never have been less than one-tenth of its present diameter.

Critics have charged that Alfvén's cosmological ideas are vague and unsupported by observations. But Timothy E. Eastman of the Space Plasma Physics Branch of the National Aeronautics and Space Administration points out that laboratory experiments with ever more powerful plasma generators and measurements by space probes have confirmed many of Alfvén's predictions concerning plasmas within the solar system, at least. "There is a revolution brewing," Eastman remarks, "in applying this knowledge to astrophysics."

Workers at the Los Alamos National Laboratory are already doing just that. With the help of supercomputers a team led by Anthony L. Peratt has created cosmological models based on recent findings about plasmas and on Alfvén's theories. One simulation shows how plasma filaments like those hypothesized by Alfvén could generate the uniform microwave background, the discovery of which was the clinching evidence for the big-bang model. Other simulations show how electromagnetic forces would help gravity to shape clouds of plasma into galaxies. The simulations, Peratt says, produce every known type of galaxy without invoking the presence of invisible "dark matter," whose gravitational effects play a crucial role in most cosmological models.

Although some plasma dissidents are also red-shift dissidents, the putative dean of the latter group, Halton C. Arp of the Mount Wilson and Las Campanas Observatories, says he does not share Alfvén's "plasma approach." Arp is concerned less with proposing alternative models of the cosmos than with undermining the Hubble relation between red shift and distance, which Arp has called the "single, frail assumption on which so much of modern astronomy and cosmology is built." Named for the American astronomer Edwin P. Hubble, who proposed it, the Hubble relation implies that objects outside our galaxy are receding from the earth at speeds proportional to their distance.

Arp says he has observed many objects with red shifts that do not conform to the Hubble relation. He maintains that quasars, for example, whose large red shifts suggest they are the most distant objects in the universe, are actually no more distant than galaxies and are probably offshoots of them. The images of most quasars, Arp points out, appear near galaxies with much smaller red shifts; moreover, luminous material appears to connect certain quasars with galaxies. Arp contends that because it is unlikely these supposedly connected objects are moving at greatly different speeds, their red shifts—and possibly all red shifts—probably result from something other than recessional velocity.

Most astronomers have dismissed Arp's observations as coincidences that statistically are not really surprising. In the past few years, however, Jack W. Sulentic of the University of Alabama has reported that his own observations corroborate Arp's claims of links between objects with different red shifts. Research in optics has given Arp's work a boost of

© 1987 SCIENTIFIC AMERICAN, INC

The best decision for millions of businesses, educators and home users.

Quality.

Reliable performance is our design objective. Our engineering team takes pride in the exceptional quality they can produce utilizing our proprietary test equipment. The result: a 33,000-hour mean time between failure.

Compatibility.

Our MS-DOS® based computers are the best-selling PC compatibles in America. Tandy 286- and 386-based business systems are OS/2™ ready as well.

Why Tandy® Computers? Because there is no better value.™

Technology.

Innovative design, custom circuitry and distinguishing features make our computers more than just clones. Plus, every Tandy desktop computer is designed and built in our own USA manufacturing plants.

Connectivity.

Tandy PC compatibles can be linked into a workgroup for communications with other PCs and mainframe computers alike.

Longevity.

Technology has been our business for sixty-six years. In 1977, we became the first company to successfully manufacture and market a personal computer—nobody's been in the business longer.

High performance, great prices and the best service and support in the industry.

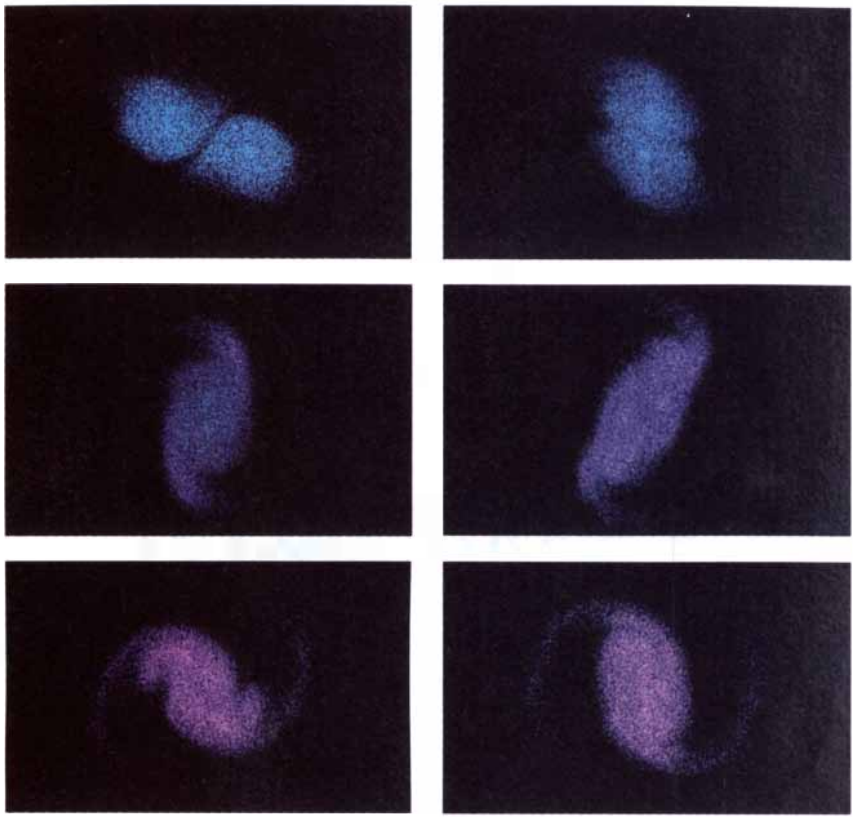
Tandy Computers:
Because there is no better value.

Radio Shack®
COMPUTER CENTERS
A DIVISION OF TANDY CORPORATION

MS-DOS/Reg. TM and OS/2/TM Microsoft Corp.

© 1987 SCIENTIFIC AMERICAN, INC

Plasma physics paints a picture of the cosmos different from that of conventional astrophysics



PLASMA CLOUD under the influence of electromagnetic forces as well as gravity coalesces into a galaxy in a computer simulation done at the Los Alamos National Laboratory.

another kind. Emil Wolf of the University of Rochester proposed last year that the spectrum of certain forms of coherent light, in which the waves travel in step with one another, can shift as the light propagates through space. This year other workers at Rochester carried out laboratory experiments—one with light and one with sound waves—that they say confirm Wolf's prediction. Wolf thinks this mechanism might be responsible for all or part of the red shift seen in the spectra of quasars.

Those astrophysicists who reject the positions of Alfvén and Arp—and most do—concede that recent observations have strained the standard big-bang model. Theorists are particularly disturbed, Jeremiah P. Ostriker of Princeton University notes, by the growing evidence of large-scale inhomogeneity in the universe's structure, which conflicts with the uniformity of the cosmic background radiation. Nevertheless, Ostriker says, theorists have proposed some solutions that are unconventional but do not require abandoning the standard model.

Lawrence M. Krauss of Yale University points out that Alfvén and Arp have attacked parts of the big-bang model without offering a comprehensive alternative. "They're working against something that explains everything very well," he says. He acknowledges, however, that the persistence of such dissident views as those of Alfvén and Arp reflects the tenuous character of all cosmological theories. "There are a lot of fundamental assumptions we base our model on," Krauss says, "that may be wrong." —*John Horgan*

All Is Flux

Nothing, including hot spots in the earth's mantle, stays still

To those who would write the history of the earth the constant flux of its crustal plates is a challenge. Although past plate motions can be reconstructed from the striped magnetic anomalies that parallel midocean spreading centers, that method does not work for the period between 115 and 80 million years ago when no magnetic field reversals took place; moreover, the stripes record only the motions of the plates with respect to one another and not their absolute motions with respect to the earth's interior. For some years now many investigators have thought volcanic hot spots, of which Hawaii is the best-known example, offered a solution.

Considerable evidence suggests that the hot spots overlie rising plumes of rock anchored deep in the mantle. If, as it has seemed reasonable to assume, the plumes are stationary, then they define a fixed reference frame, and the tracks they leave at the surface—island chains and undersea volcanoes, for example—record the absolute motions of the plates.

Yet hot spots are not fixed, according to a report in *Nature* by Peter Molnar and Joann Stock of the Massachusetts Institute of Technology. The workers started their analysis with the Hawaiian hot spot, whose track on the Pacific plate (beginning with the Emperor Seamounts and continuing with the Hawaiian Islands) is well defined and well dated for the past 68 million years. They then used relative plate motions derived from magnetic anomalies to calculate where the tracks of hot spots should lie on other plates, assuming that the plumes themselves are not in motion. By comparing these predicted tracks with the observed tracks of five hot spots—those associated with the volcanic islands of Iceland, Tristan da Cunha in the South Atlantic and Réunion, Kerguelen and St. Paul's in the Indian Ocean—Molnar and Stock put the fixity hypothesis to the test.

In all five cases the workers found that the calculated hot-spot tracks diverged from the observed tracks by hundreds of kilometers. The results imply that the underlying plumes must be moving with respect to the Hawaiian plume at a rate of between one centimeter and two centimeters per year, or roughly one-tenth as fast as the plates. The conclusion does not mean that all reconstructions of plate motions based on hot spots are invalid, but it does place a limit on their precision. According to Molnar, investigators will have to be more cautious in reconstructing comparatively small-scale geologic events,

© 1987 SCIENTIFIC AMERICAN, INC

EVERY month Eastern Airlines has five days to assign 12,000 pilots and flight attendants to thousands of different routes. But flight crews aren't automatically assigned a monthly world schedule. They bid on routes, and assignments are made on the basis of seniority.

For years this process was handled manually. Eastern had to deliver route selections to its employees, wait for employees to fill out their bids and return them to a specific destination, put route preferences in order of seniority, make assignments and return approved schedules to employees.

The whole process was unwieldy and inefficient. And Eastern wanted an automated solution that would relieve this aggravation.

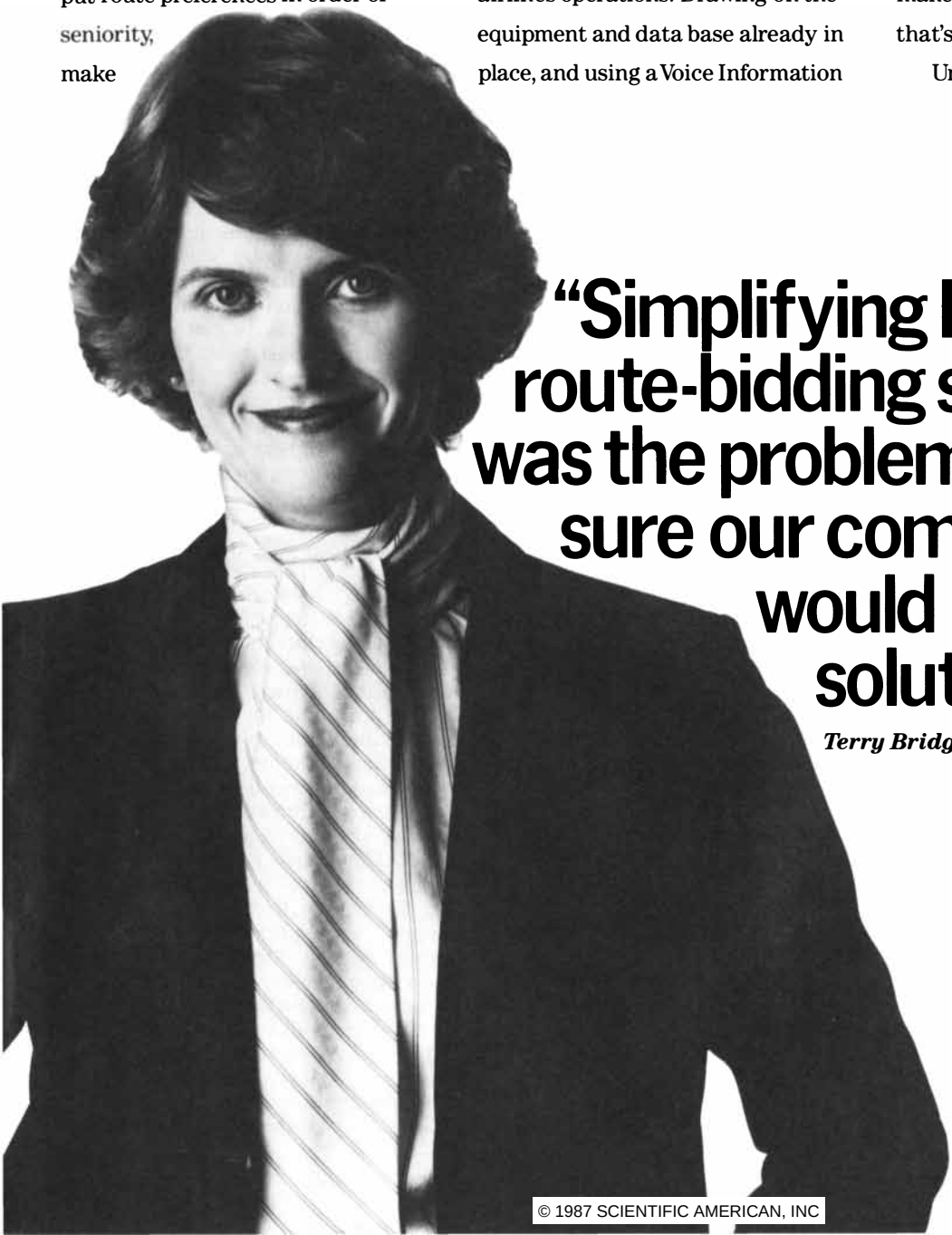
So Terry Bridges and Unisys were called into action. "Eastern had 12,000 employees traveling throughout the western hemisphere and no easy way to reach them."

They were already using Unisys computers to handle other aspects of airlines operations. Drawing on the equipment and data base already in place, and using a Voice Information Processing System in an innovative way, they found the solution they were looking for. Eastern and Terry developed the first automated crew bidding system in the world.

"Now, flight crews just pick up a touchtone phone and they're into the system. Going step-by-step through the interactive system, they can enter their bids, receive their bid awards and even swap flight assignments.

"We all really worked together—crews, management and Unisys—to make everybody's life easier. And, well, that's the power of 2."

Unisys and airlines. The power of 2.



"Simplifying Eastern's route-bidding system was the problem. We were sure our computers would be the solution."

Terry Bridges, Account Executive, Unisys

UNISYS

The power of 2

© 1987 SCIENTIFIC AMERICAN, INC

PUT STARS IN HER EYES

with a Questar, for it is a gift that will go on giving. If she is that rare child with a true curiosity, it will help her to discover the nature of things all her life. It will be her own personal contact with the universe, which she will have the excitement of seeing, from her very first glimpse, through the finest of telescopes. Both optically and mechanically, Questar will respond rewardingly to her touch so that she will never know the frustration of an inferior tool. Questar's beauty and versatility will make it a cherished possession and an invaluable companion, whether her main interest becomes the study of the sun, moon, and planets, or an earth-oriented observing of nature close at hand. From her first-hand investigation will come an expanding awareness and continuing enjoyment. And its easy portability will permit her to take it with her wherever she travels.

To open to such a young mind that vast store of knowledge to which she may one day make her own contribution, will be your rare privilege.

© Questar Corporation 1984

The Standard Questar with its beautiful star chart is shown above; at right is the Duplex Questar in the leather carrying case which complements both models.

LET US SEND YOU OUR LITERATURE DESCRIBING QUESTAR, THE WORLD'S FINEST MOST VERSATILE TELESCOPE. PLEASE SEND $2 TO COVER MAILING COSTS ON THIS CONTINENT. BY AIR TO SOUTH AMERICA, $3.50; EUROPE AND NORTH AFRICA, $4; ELSEWHERE $4.50; INQUIRE ABOUT OUR EXTENDED PAYMENT PLAN.

QUESTAR

Box 59, Dept. 80 P, New Hope, PA 18938
(215) 862-5277

such as the collision of small terranes with western North America.

Whereas the notion of fixed hot spots has been convenient for paleogeographers, it has troubled workers who study the dynamics of the mantle, according to Peter Olson of Johns Hopkins University. Writing in the same issue of *Nature,* Olson argues that the hot-spot drift detected by Molnar and Stock makes physical sense. Recent seismic results suggest that the lower mantle is in vigorous convective motion, driven by heat flowing out of the core. Since the plumes are thought to originate near the core-mantle boundary (at a depth of about 2,900 kilometers), it would be surprising if they were not moving too. *—Robert Kunzig*

Currents in Plastics

Workers make an organic polymer whose conductance rivals copper's

European investigators in search of better conductive materials have been extraordinarily fecund of late. It was Swiss workers who discovered last year that certain ceramic oxides can superconduct at record-high temperatures (about 100 degrees Kelvin at last count). Now investigators at the BASF Laboratories in Ludwigshafen, West Germany, have refined an organic polymer that displays the highest conductivity at room temperature, for its weight, of any known material.

Although the new organic polymer is overshadowed by the astounding advances in superconductors, it too represents a breakthrough in solid-state physics, according to Alan J. Heeger of the University of California at Santa Barbara. Heeger, who has tested the polymer, thinks it may be the prototype for a long-sought plastic conductor—one that could be much cheaper, lighter and more malleable than metal conductors.

The base of the conductor developed at BASF by Herbert Naarmann and N. Theophilou is polyacetylene, which consists of chains of carbon and hydrogen atoms. When it is "doped" with iodine according to a special process, the polyacetylene has a conductance of about 150,000 siemens (a siemen is the reciprocal of an ohm, the unit of resistance) per centimeter, Naarmann and Theophilou report. That is about a fourth the conductance of an equal volume of copper. By weight the polymer is

© 1987 SCIENTIFIC AMERICAN, INC

twice as conductive as copper, making it the world's most conductive material, according to Arthur J. Epstein of Ohio State University.

Heeger and Alan G. MacDiarmid of the University of Pennsylvania discovered 10 years ago that iodine doping makes polyacetylene a conductor. Polyacetylene is ordinarily a semiconductor; the addition of iodine apparently pulls tightly bound electrons out of the polymer chains, leaving positively charged holes that make conduction possible. The samples produced by Heeger and MacDiarmid, however, had a conductance of only about 200 siemens per centimeter. They also decomposed in the open air within a few hours. The BASF material, in contrast, can maintain its molecular stability in the open air at room temperature for more than a month.

The material's improved conductance and stability both seem to derive from BASF's synthesizing process, which leaves the polymer's molecules considerably less jumbled and fragmented than usual. Epstein suggests that electrons can move with less interference along the longer, straighter molecules and that such molecules are less susceptible to oxidation, the main cause of instability. The fact that the polymer's conductance does not increase at lower temperatures, Heeger says, suggests it is still somewhat impure and hence not as stable or conductive as it could be. Indeed, there are informal reports that newer samples produced at BASF have a conductance higher than copper even by volume.

In its present form the BASF material is not appropriate for widespread application, because it lacks the major qualities that make plastics attractive. Polyacetylene does not dissolve in water or any other liquid, including acids. Moreover, it does not melt or become pliable when heated; it decomposes. Heeger thinks more research should turn up related polymers that are both highly conductive and processible by conventional means. "There is no reason to believe [high conductance] is specific to polyacetylene," he says. —*J.H.*

Another Sell-Out From the U.S. Mint?

This history-making Silver Eagle Coin (shown actual size) is the largest, heaviest, *purest* silver coin the U.S. government ever minted. First issued in October, 1986, it became an instant rarity, so that demand from collectors and investors has already doubled it's value.

Now's the time to invest in new 1987 U.S. Silver Eagle coins—*before* history repeats itself—for they are already appreciating in value. In Brilliant Uncirculated condition, only $19 each, plus $1 p. & h., per coin. U.S. Mint tube of 20 only $350, plus $20 p.&h. per tube. 10 day money back guarantee.

TO ORDER: Call toll free, including Sat. & Sun., 1-800-USA-5858, and charge to Visa, Mastercard, or Amex. Or send check or money order to Federal Coin Fund, Dept. CC03 3 North New St., Dover, DE 19901.

Limited quantities available.

TECHNOLOGY

Plain Sense

A prairie experiment aims to make more sense of satellite imagery

Satellites provide a powerful tool for monitoring droughts, deforestation, crop fluctuations and other natural and man-induced processes on a broad scale. But although the resolution of satellite-based instruments has increased, the ability of scientists to interpret the images has not kept pace.

This year more than 100 workers from the U.S., Canada and Great Britain are engaged in an unusual experiment on the plains of Kansas to correct the situation. The investigators, who include biologists, meteorologists and remote-sensing experts, are observing a tract of land and the overlying atmosphere directly (at ground level and from planes and helicopters) while remote-sensing satellites pass above. Over the next two years the workers will compare the satellite images with the other data in order to determine how the interpretation of such images can be improved.

The focus of the $10-million experiment, funded mainly by the National Aeronautics and Space Administration, is a square of gently rolling prairie, cut by an occasional gully, measuring 15 kilometers on a side. The land was chosen in part for its proximity to Kansas State University in Manhattan, a center of research in remote sensing, and also because "we wanted something fairly representative," according to Forrest Hall of NASA's Goddard Space Flight Center.

An important goal of the project, Hall says, is to understand "how data scales from the ground level to the remote-sensing level." All told, five satellites are gathering images of the site, at altitudes ranging from 40,000 kilometers to about 750 kilometers. Meanwhile a U-2 is scanning the area from high in the stratosphere and six planes and a helicopter are taking readings at lower altitudes. Instruments on the ground measure precipitation, wind speed, the moisture content of the soil and other variables, and workers collect samples of vegetation.

Ghassem Asrar of Kansas State notes that the project should show how heat, humidity and other atmospheric conditions can distort satellite images from moment to moment; this knowledge, he says, may suggest ways to eliminate the distor-

We've got a job so important, it takes a retired person to do it.

We want you to help give new lives to the 27 million Americans who are functionally illiterate.

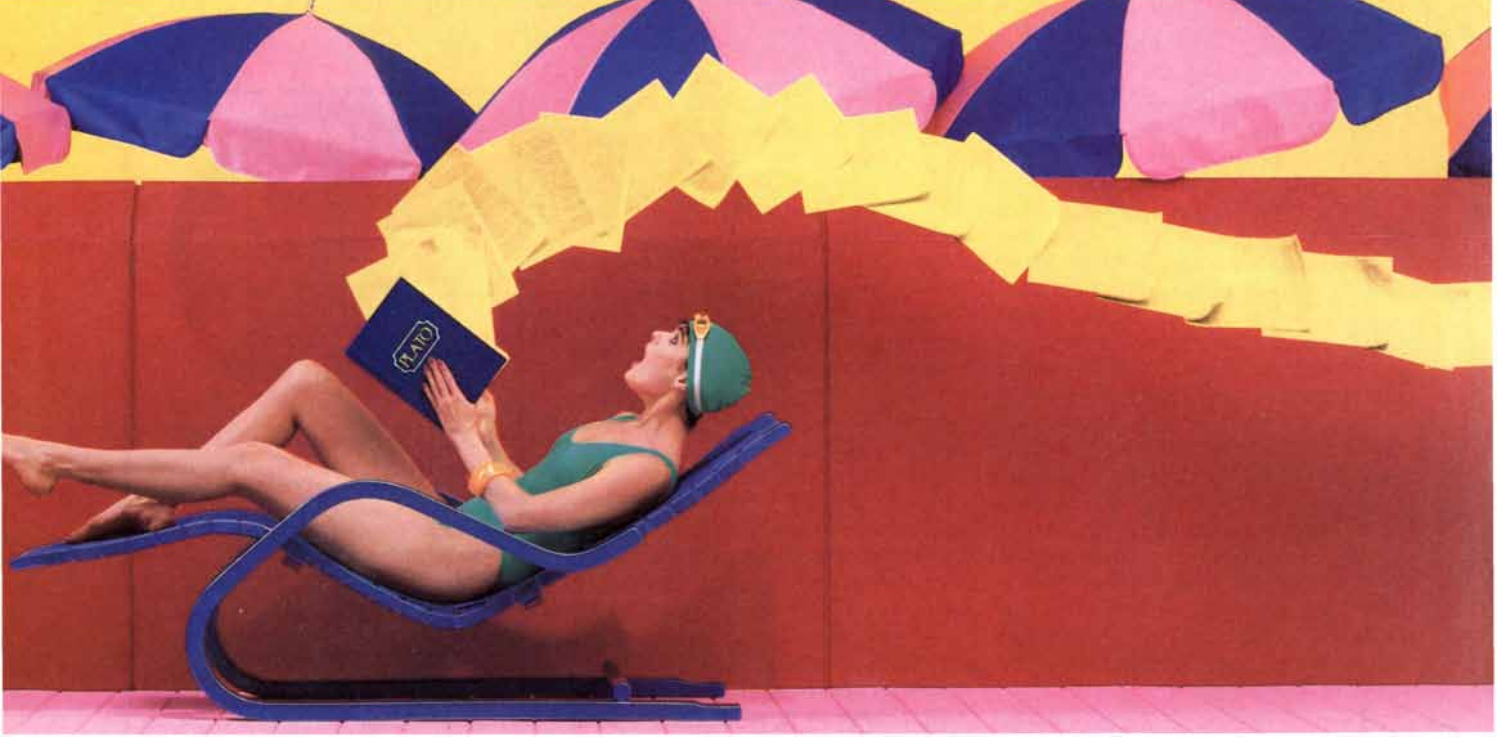
Today, illiteracy has become an epidemic that has reached one out of five adults. But you can help stop it by joining the fight against illiteracy.

Call the Coalition for Literacy at toll-free **1-800-228-8813** and volunteer.

Volunteer Against Illiteracy. The only degree you need is a degree of caring.

Ad Council Coalition for Literacy

© 1987 SCIENTIFIC AMERICAN, INC



NOTHING ATTRACTS LIKE THE IMP

CORIANDER SEEDS FROM MOROCCO ANGELICA ROOT FROM SAXONY JUNIPER BERRIES FROM ITALY CASSIA BARK FROM INDOCHINA

tion automatically. The experiment should also unveil the complex processes that can alter a satellite's view of the earth's surface from day to day and season to season: changes in the reflectivity of the soil and in the color and density of the vegetation, for example. "You have three dynamical systems: soil, plant life and the atmosphere," Asrar says. "Any change in one affects the others." —*J.H.*

Butterfly Catcher

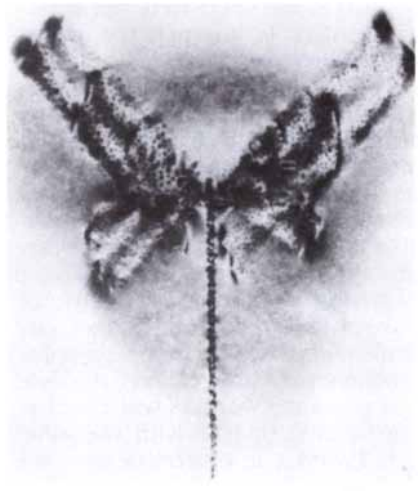
Butterfly-shaped flaws gather up metallic impurities in silicon

Metallic impurities in a crystal of silicon degrade the performance of integrated circuits laid out in the crystal's surface layer. Such impurities are usually removed by heat-treating the crystal so that another impurity, oxygen, can combine with them to form precipitates. The precipitates migrate away from the surface and also create internal structural defects that attract remaining contaminants. Employing one impurity in a semiconductor to get rid of others is called gettering.

A group of investigators at the Massachusetts Institute of Technology has now found a new gettering process for crystalline silicon. Like the standard process, it depends on heat to form precipitates and structural defects, but it can be applied to oxygen-free silicon and can probably be made to work much more quickly. The special charm of the process is that under an electron microscope the defects it creates look remarkably like butterflies.

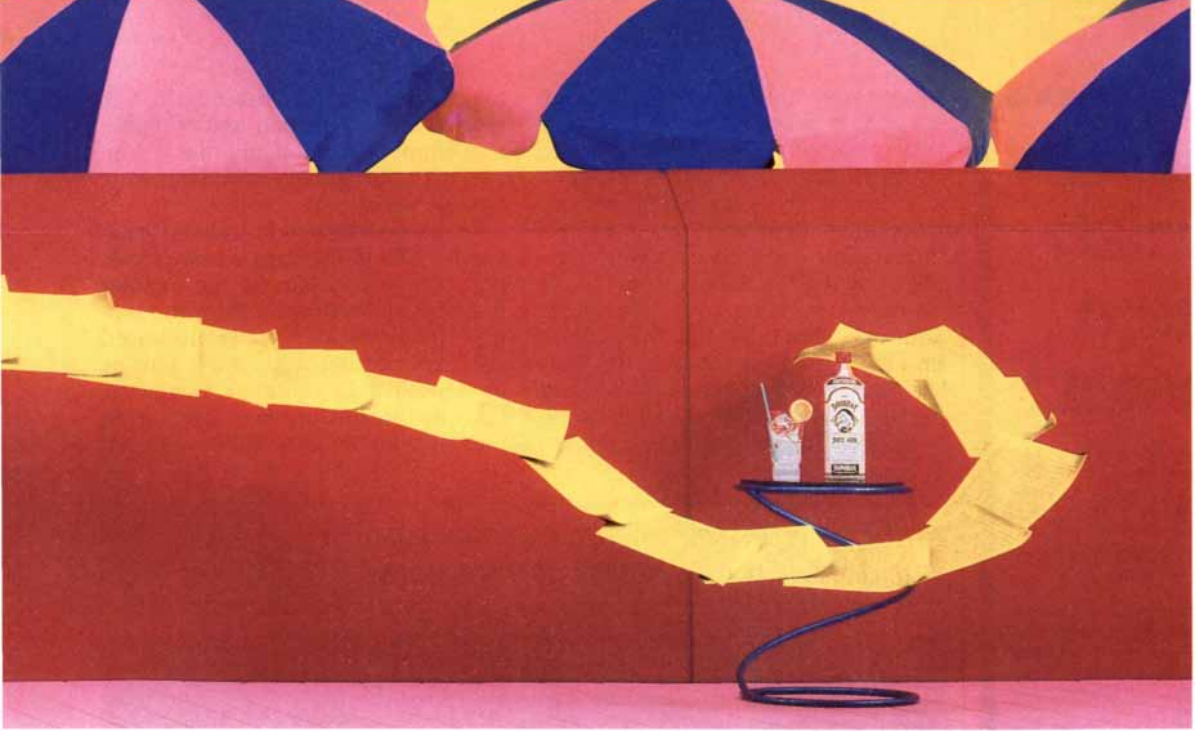
In the new gettering process, developed and studied in the laboratory of Harry C. Gatos and Jacek Lagowski, the silicon crystal is heated to a temperature of 1,170 degrees Celsius for six hours in an atmosphere of oxygen and hydrochloric acid. It is then placed in nitrogen and heated to 700 degrees for 20 hours and to 1,000 degrees for six more hours. A crystal subjected to the process develops "butterfly defects," which were first glimpsed by Osamu Ueda, a visiting scientist from Fujitsu Laboratories, Ltd., in Japan, working with graduate students Krys Nauka and Mark Goorsky.

Lagowski theorizes that the defects evolve from interstitial atoms of silicon: silicon atoms that are wedged between the regular planes of the crystalline lattice. In the course of the new gettering process the interstitial silicon atoms are removed from the surface layer, leaving most of the interstitial atoms in the underlying material. There they aggregate and grow into complexes that cause deformations in the regular atomic lattice, known as loop dislocations. At the same time metallic impurities interact with the complexes to form



BUTTERFLY is actually a microscopic tangle of flaws and precipitates within a crystal of silicon.

© 1987 SCIENTIFIC AMERICAN, INC



precipitates, which also result in loop dislocations. The proliferation of interlocked dislocations distorts the surrounding array of atoms, allowing metallic atoms to approach and be captured by the dislocations as the temperature is raised in the final step of the process. The entangled dislocations and precipitates actually form a complicated three-dimensional structure about a micrometer in diameter that looks like a butterfly only when it is seen from a particular angle.

Although the butterfly defects took shape during more than 30 hours of heat treatment, the investigators are confident that the defects can be formed much more quickly—perhaps in seconds rather than hours. They are working with Texas Instruments, Inc., on an exceedingly rapid gettering process based on fast-growing butterflies. *—Gregory Greenwell*

BIOLOGICAL SCIENCES

High-Tech Horses

Can biotechnology give a lift to American jumpers?

The equestrian sport of jumping, once ruled by European riders, has recently been dominated by Americans. Indeed, the U.S. team won the gold medal in show jumping in the 1984 Olympics. Almost all the horses these new competitors ride, however, are still European, the product of centuries of careful breeding.

Some U.S. breeders are counting on biotechnology for a quick remedy. By bringing embryos from a champion mare to term in surrogate mothers, thereby increasing her productive output to 15 or more foals per year, they hope to compress considerably the period of trial and error needed to create a new breed of homegrown "sport horses."

Colorado State University has led the effort to refine modern breeding technologies and make them available to raisers of sport horses. B. W. Pickett, who heads the university's equine sciences program, notes that although cattle growers have exploited biotechnology for decades, it has attracted interest within the more conservative, tradition-bound realm of horse breeding only in the past few years.

A major reason, Pickett says, is that racing has long been the dominant equine sport in America, and the organizations that oversee the racing breeds—the thoroughbreds, standardbreds and quarter horses—prohibit or severely limit the use of breeding technologies even as pedestrian as artificial insemination. To compete in international jumping events, in contrast, a horse need not belong to any specific breed. The growing popularity of these events, Pickett notes, combined with an increasing amount of prize money, has made proved winners—and their offspring—increasingly valuable.

To carry out an embryo transfer, workers at Colorado State artificially inseminate a mare and one week later flush the uterus with a saline solution. If fertilization has occurred, the solution will flush out an embryo about a millimeter in diameter, consisting of from 60 to 90 cells still encased in the egg's outer membrane, or shell. After locating the embryo under a microscope, the workers transfer it to the uterus of a surrogate mare either through an incision in

© 1987 SCIENTIFIC AMERICAN, INC

¡Speak Spanish like a diplomat!®

What sort of people need to learn a foreign language as quickly and effectively as possible? *Foreign service personnel,* that's who. Members of America's diplomatic corps are assigned to U.S. embassies abroad, where they must be able to converse fluently in every situation.

Now you can learn to speak Spanish just as these diplomatic personnel do — with the Foreign Service Institute's Programmatic Spanish Courses. You'll learn Latin American Spanish recorded by native speakers.

The U.S. Department of State has spent thousands of dollars developing this course. It's by far *the most effective way* to learn Spanish at your own convenience and at your own pace.

The course consists of a series of cassettes and accompanying textbook. Simply follow the spoken and written instructions, listening and repeating. By the end of the course, you'll be learning and speaking entirely in Spanish!

This course turns your cassette player into a "teaching machine." With its unique "programmatic" learning method, you set your own pace — testing yourself, correcting errors, reinforcing accurate responses.

The FSI's Programmatic Spanish Course comes in two volumes, each shipped in a handsome library binder. Order either, or save 10% by ordering both:

☐ **Volume I: Basic.** 12 cassettes (17 hr.), manual, and 464-p. text, $135.
☐ **Volume II: Intermediate.** 8 cassettes (12 hr.), manual, and 614-p. text, $120.

(CT residents add sales tax)

TO ORDER BY PHONE, PLEASE CALL TOLL-FREE NUMBER: **1-800-243-1234.**

To order by mail, clip this ad and send with your name and address, and a check or money order–or charge to your credit card (AmEx, VISA, MasterCard, Diners) by enclosing card number, expiration date, and your signature.

The Foreign Service Institute's Spanish course is unconditionally guaranteed. Try it for three weeks. If you're not convinced it's the fastest, easiest, most painless way to learn Spanish, return it and we'll refund every penny you paid. Order today!

130 courses in 46 other languages also available. Write for free catalog. Our 15th year.

Audio-Forum
Room K926
On-the-Green,
Guilford, CT. 06437
(203) 453-9794

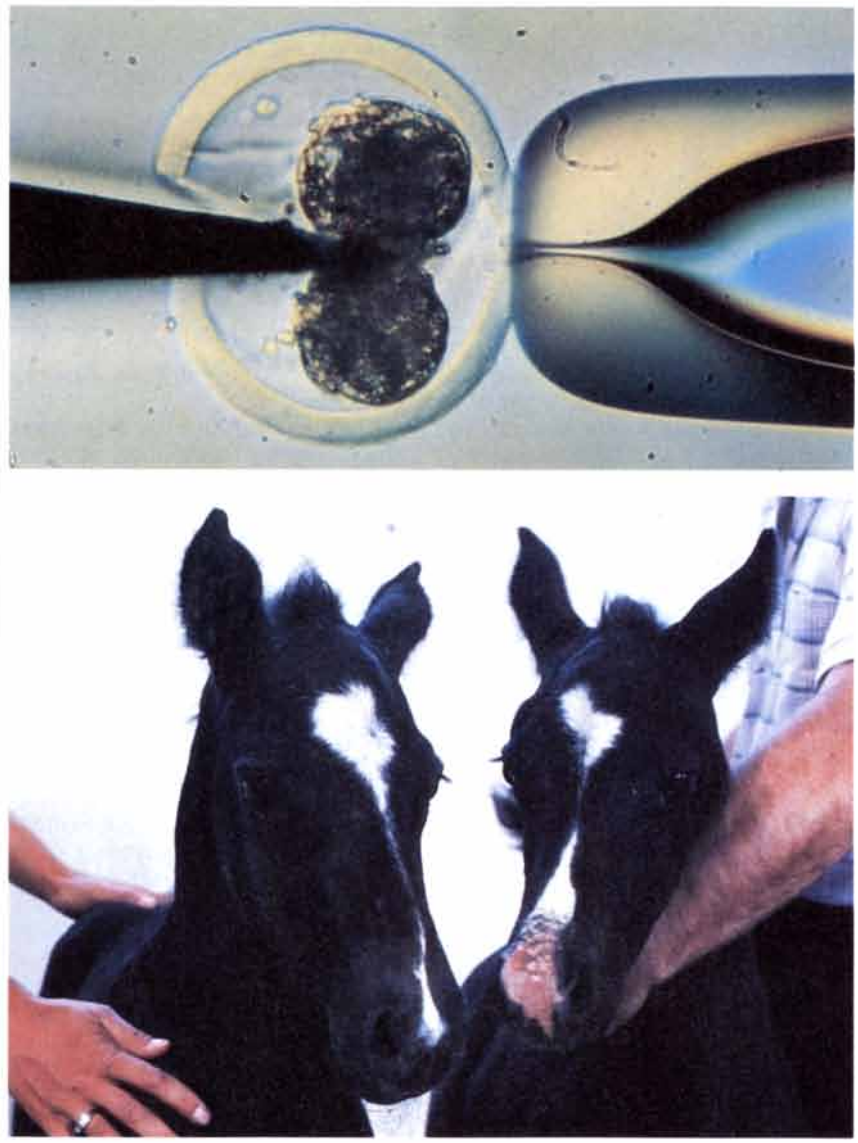
AUDIO-FORUM®

her side or through her birth canal. The donor mare is then free to conceive another embryo or to return to competition.

In the past five years embryo transfers done at Colorado State have produced some 150 foals from stock supplied by half a dozen sport-horse breeders, according to Edward L. Squires, who oversees the equine breeding program. Many of the foals are crosses between European warm bloods, even-tempered and relatively thickset horses that excel at jumping, and American thoroughbreds, known for their speed and competitive temperament—a mix that is considered a particularly promising one, Squires says.

Before an embryo is transferred it can be split to increase a mare's output even more. Part of the cellular material is withdrawn from the embryo and placed in the hollowed shell of another egg; each part is implanted in a different surrogate, thereby producing two foals from one embryo. Embryo splitting is still

Embryo transfer produces foals faster, but does it make better competitors?



ONE-WEEK-OLD EMBRYO flushed from an artificially inseminated mare is cut in two while held by suction from a pipette (top)*; each half can then be implanted in a surrogate mother. The technique served to produce the genetically identical twins Question* (with the sunburned nose) *and Answer at Colorado State University three years ago.*

© 1987 SCIENTIFIC AMERICAN, INC

experimental, but a potential spinoff of the technique is already evident: the genetically identical foals the method yields would enable breeders to evaluate the effects of different training or nutritional programs with great precision.

Peter H. Birdsall, a Canadian veterinarian and authority on sport-horse bloodlines who in 1984 formed the Horses in Sport Federation to keep track of the budding breed, believes that if the methods are judiciously applied to high-quality stock, they could help U.S. breeders catch up to the Europeans. Otherwise, he warns, the techniques could quickly flood the market with mediocre horses.

As for those sport horses already produced by embryo transfer, Birdsall says it is "too early to tell" whether they will live up to expectations: the first crop of five-year-olds is only now ready for competition. *—J.H.*

Potato Eaters

Mother was wrong: the skin may not be "the best part"

Two groups at Cornell University have separately reported some good news and some bad news about potatoes. One group has come up with a hybrid potato that may be much more resistant to insects than those now grown in North America. The other group warns that potatoes cooked in their skins may sometimes be harmful.

First the good news. In the 1970's Ward Tingey and Robert Plaisted read about a wild Bolivian potato, *Solanum berthaultii,* that has what seemed to be a potentially valuable trait. The species exudes a clear chemical from microscopic glandular hairs on its leaves. On contact with air this mixture of phenolic compounds and polyphenyl oxidase results in a brown, tarlike glue that sticks to the feet and mouthparts of soft-bodied insects such as aphids and traps them. Larger insects such as the Colorado potato beetle (the most serious U.S. pest of potatoes) find it difficult to get a foothold on leaves covered with the tar; it is therefore hard for them to feed on the leaves, and so they mature and reproduce more slowly.

For the past decade Tingey and Plaisted have sought to transfer this insect-resistance trait to a species adapted to North American conditions. The team has now developed hybrids estimated to have a resistance to insects equal to that produced by three or four applications of insecticide. The hybrids have also inherited some undesirable traits from their Brazilian progenitor. For example, they stop producing tubers when the day is more than 12 hours long. Tingey thinks that in a decade or so these traits can be eliminated, and that farmers will have a tougher plant with a higher yield of potatoes.

Now the bad news: potatoes can make you sick, according to Nell I. Mondy and Barry T. Gosselin of Cornell. Concentrated in the skins of potatoes are toxins called glycoalkaloids, which are produced in response to stresses such as insects or heat. Peeling potatoes before cooking gets rid of glycoalkaloids; Mondy and Gosselin maintain that boiling potatoes in their skins or frying the skins can heighten the problem. The toxins apparently migrate into the potato's flesh when it is boiled; frying the skins eliminates water and concentrates the toxins.

Eating too large a quantity of the glycoalkaloids results in headaches, nausea and diarrhea. People do not often associate such symptoms with the eating of potatoes, Gosselin says, but there are documented cases of illness and—very rarely—death. In 1979, for instance, 17 English schoolboys were hospitalized after eating toxic potatoes. The latest documented death occurred in a village in Cyprus in 1933. The level of toxins in a potato's peel depends on the variety, the conditions of the growing season and the method of storage. A single potato might give rise to symptoms if it has turned green under display lights or has sprouted in an unventilated storage bin.

What about societies that have depended on potatoes, such as poor Europeans in the 18th and 19th centuries and some South American Indians? The Irish most likely peeled theirs, Mondy says. And there is evidence that the Incas, in the potato's land of origin, stored the tubers in cool, dry places. *—Elizabeth Collins*

Beeboppers

U.S. bee experts disagree over where, or whether, to stop "killer bees"

The saga of the Africanized bees began in the 1950's. A geneticist brought African honeybees to Brazil, hoping to breed better honey producers than the indigenous variety, an earlier import from Europe. The African bees escaped, and they soon proved to be less adept at producing honey than at viciously defending their hives and "Africanizing," or genetically overwhelming, the Europeans. Over the next three decades the Africanized bees spread through South and Central America, earning the sobriquet "killer bees" for their swarming, occasionally fatal attacks on animals and humans.

AFRICANIZED BEES surround a single European bee (top center) *in a photograph made by Orley R. Taylor, Jr. European bees are slightly larger; otherwise the two strains are virtually indistinguishable.*

They are now in southern Mexico, moving north at about 200 miles per year. Most—but not all—U.S. bee experts agree that Africanized bees pose a genuine threat, less to human lives or even to the honey industry than to the crops (worth billions of dollars) that European honeybees pollinate during managed releases; beekeepers, these authorities say, would have great difficulty managing Africanized bees. The authorities strongly disagree, however, about where and how to stop them.

The U.S. Agricultural Research Service (ARS) wants to choke the bees off at Mexico's Isthmus of Tehuantepec, just north of where they were last spotted. The service originally sought congressional approval for a $10-million project that would span the entire isthmus. U.S. and Mexican workers would destroy nests of Africanized bees, kill their drones with pheromone-baited traps and replace their queens with European queens, all the while keeping the European hives isolated. Officials would also quarantine the area, checking vehicles that might carry bees north.

Other bee investigators charged that the project would be unwork-

© 1987 SCIENTIFIC AMERICAN, INC

able and wasteful, and this past May the ARS presented a much less ambitious plan involving only the Pacific side of the isthmus. ARS workers still insist that even the scaled-down, $2-million project, which they expect to begin this fall, can stop or at least stall the bees. "The bees are not challenging the Eastern side of the isthmus, and so the plan is to trap them on the other side," Hachiro Shimanuki of the ARS says. "Even if we fail, we'll learn."

Orley R. Taylor, Jr., of the University of Kansas says he supports the ARS efforts to develop "bee suppression" techniques but calls their goal of stopping the bees a fantasy. "Mexico is a very big country and this bee is a supertramp," he says. "They don't have the money, the personnel or the technology to stop it."

Taylor is seeking funds to carry out a study in Jalapa, about 200 miles north of the isthmus, where he is planning to monitor the "genetic swamping" of European bees by Africanized ones. He is also trying to cultivate European strains that are less likely to become Africanized. He has isolated some European bees, for example, whose mating flights tend to take place at times of day different from those of the Africanized bees. Such research, perhaps combined with what is learned from the ARS project, Taylor says, may lead to an effective program to stop Africanized bees in the future.

But should the bees be stopped? Roger A. Morse of Cornell University says his studies of Africanized bees in Africa and South America have led him to believe they are good honey producers and pollinators; best of all, they are resistant to diseases that currently plague European bees. He concedes they are "nasty little bees" but maintains they are manageable and may actually be a boon for U.S. agriculture. In fact, Morse says, he suspects U.S. farmers have already begun importing Africanized bees to pollinate their crops. —*J.H.*

MEDICINE

Never a Right Age

But the divorce of parents is hardest on adolescents

Of the 1.2 million children in the U.S. who experience divorce every year, teen-agers are often assumed to be the ones most capable of coping with the disruption of their homes. Actually, though, the impact of divorce on the already "tumultuous passage" through adolescence can cause lasting psychic and social injury, Murray M. Kappelman of the University of Maryland School of Medicine writes in *American Family Physician*. There is therefore no reason for parents heading toward divorce to wait until they think their children are "old enough to handle it."

Adolescents depend on parental role models in trying to understand the behaviors and attitudes of each sex, according to Kappelman. Teen-agers who must do without either role model at a time of sexual maturation may experience a deep sense of abandonment and become suspicious of emotional relationships. Adolescence is also a time of open or tacit rebellion. A teen-ager's obstreperous attempts to establish an independent self-identity may make it appear that the preservation of family structure is less critical for him or her than it is for a younger child, Judith S. Wallerstein of the University of California at Berkeley remarks, but in fact it is precisely in this testing period that a safety net is most needed.

Kappelman points out that teen-agers often blame themselves for the rift in the family. Moreover, they may be thrust by divorce into responsibilities and concerns beyond their years. Particularly if the adolescent is the oldest child in the home, he or she may be forced to take on the role of the absent parent or to support the custodial parent emotionally. Then too, the state of the family's finances often evokes fear: How will we live? Can I go to college? Such concerns are well founded. According to Bureau of the Census data, more than 80 percent of the children whose parents have been divorced live with their mothers; the 1985 median income for such single-parent families with children from the age of 12 through 17 was $15,249 (compared with $27,600 for families that had a father at the head).

Although there is "never a right age" for divorce, Peter A. Wish and Leslie-Beth Berger of the New England Institute of Family Relations think an easier time for both parent and child may be when the child is between the ages of 9½ and 12, although such children do often respond to divorce by doing poorly in school. Still younger children, on the other hand, may become dependent and demanding.

Teen-agers, however, tend to act out their sense of rejection and displacement, and so their response is often the most difficult to handle and the most dangerous. As Kappelman points out, they are more likely than younger children to have access to such dangerous modes of response as drugs, alcohol and suicide. —*E.C.*

Bespoke Bactericide

Compounds tailored to fool bacteria may augment the antibiotic arsenal

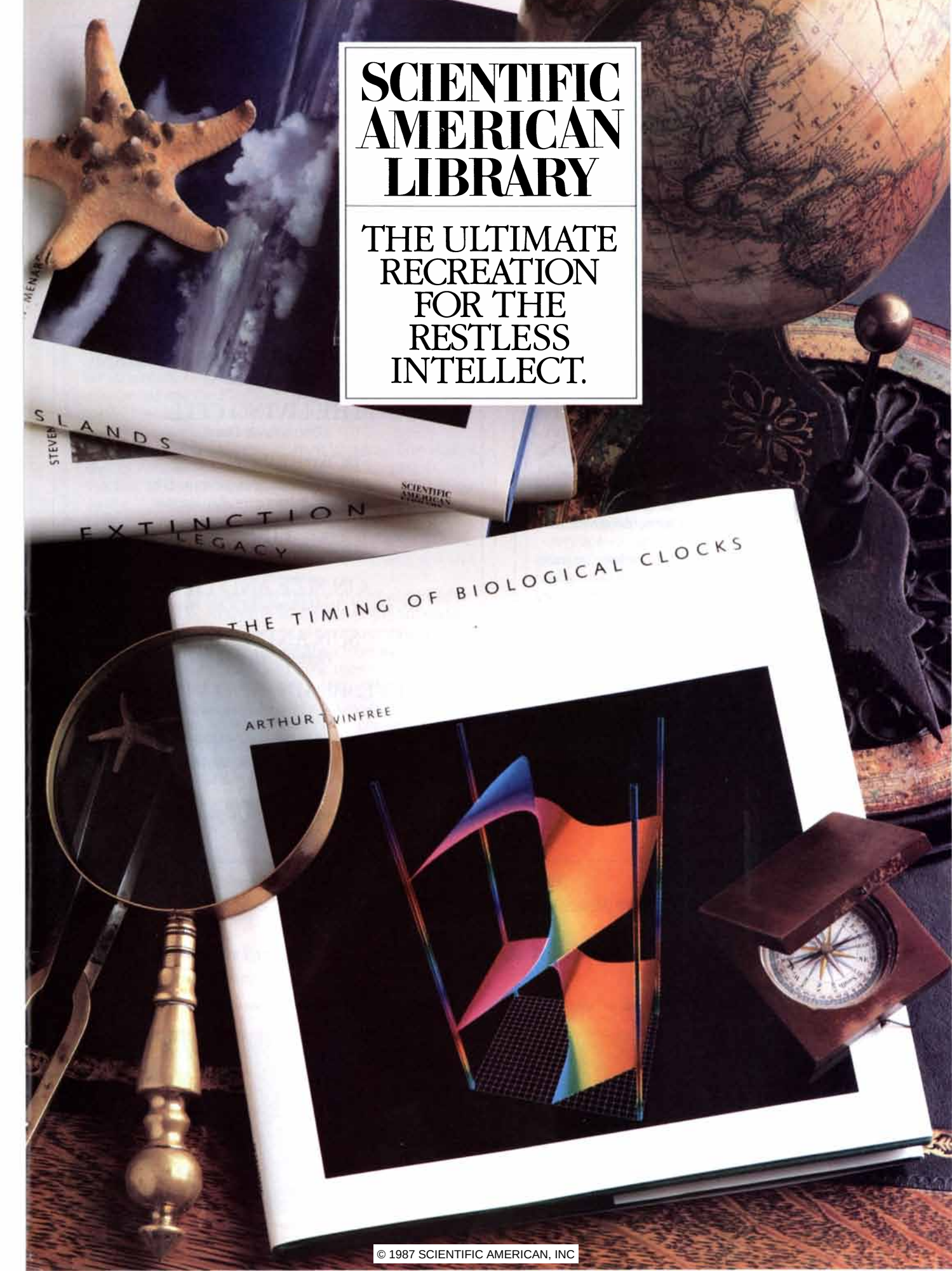
Although the discovery of antibiotics led to great advances in clinical practice, it did not provide a panacea for bacterial illnesses. There are some types of bacteria against which antibiotics are ineffective. Furthermore, some strains have developed resistance to all major antibiotics. The answer to both problems may lie in tailoring drugs to achieve specific bacteria-killing effects, as a Swedish research group reports in *Nature*.

The group in Sweden, headed by Stephen M. Hammond of Astra Alab, chose as their target the gram-negative bacteria (so called because they fail to absorb a stain devised by the 19th-century Danish physician Hans Christian Joachim Gram). Among the gram-negative bacteria are gonococci, which cause gonorrhea, meningococci, which cause meningitis, and the various forms of salmonella, a major gastrointestinal pathogen. Although there is a need for agents that act only on gram-negative types, few such compounds exist, the investigators note.

In their attempt to formulate such a drug, the workers in Sweden chose to focus on the synthesis of the membrane that encloses the bacterium. A significant component of the membrane in gram-negative bacteria is a molecule called KDO. Hammond and his colleagues found that an analogue of KDO called NHdKDO interrupts membrane synthesis. The reason seems to be that the analogue resembles KDO closely enough to be taken up by assembly enzymes but lacks the chemical structure necessary for contributing to the growing membrane.

Yet a problem remained: NHdKDO

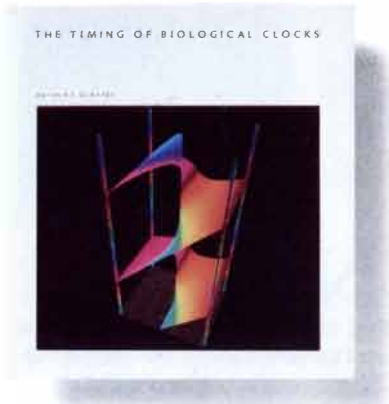
© 1987 SCIENTIFIC AMERICAN, INC



© 1987 SCIENTIFIC AMERICAN, INC

THE TIMING OF BIOLOGICAL CLOCKS

Arthur T. Winfree



In 1931, Wiley Post flew around the world in eight days and became the first human being to experience jet lag. The fatigue and disorientation he felt happens to all of us when rapid long-distance travel knocks our "internal clocks"—our circadian rhythms—out of kilter with local time. Arthur Winfree describes jet lag as "that disconcerting sensation of time travelers that their organs are strewn across a dozen time zones while their empty skins forge boldly into the future."

In the early 1930s, Wiley Post used the Winnie Mae to study jet lag and its affect on pilot performance. Courtesy of the National Air & Space Museum, Smithsonian Institution.

Jet lag, biorhythms, mosquito insomnia, temporal isolation experiments, the sleep movement of plants, "forbidden phases" of sleep when one cannot awake spontaneously—these are some of the many fascinating aspects of circadian rhythms that Dr. Winfree explores in *The Timing of Biological Clocks.*

Winfree shows that the most critical property of biological clocks is their ability to be reset on cue, enabling them to regain synchrony with a changing environment (as when we travel across time zones) or to adjust the body's 25-hour rhythm to the 24-hour solar day.

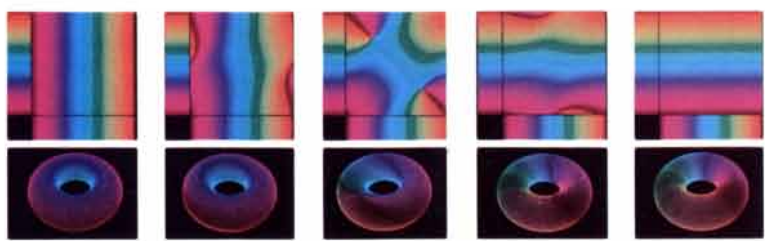
Flowering in the morning glory, as in many other plants, is timed by a circadian clock. Courtesy of Travis Amos.

Reporting experiments on animals, plants, and single cells, he not only illustrates the principles that guide the resetting of biological clocks but reveals that each of these clocks has a vulnerable phase when a suitably intense cueing stimulus can produce a thoroughly unpredictable resetting—perhaps even annihilating the clock's rhythm entirely.

The graphics that Winfree uses are as innovative as his insights. By using gradient color rather than the conventional clock dial to express the passage of time, Winfree helps us visualize the true continuities—and discontinuities—of the internal cycles that govern the processes of life.

Has this cave salamander, living in temporal isolation, lost its circadian rhythmicity? Courtesy of Chip Clark.

Arthur T. Winfree is one of the world's foremost theoreticians of circadian rhythms. Trained as a biophysicist, Winfree received a MacArthur Grant for his work on biological clocks. Formerly at the Institute for Nonlinear Science at the University of California at San Diego, he is now with the Department of Ecology and Evolution at the University of Arizona at Tucson. Professor Winfree is also the author of the classic work *The Geometry of Biological Time* and *When Time Breaks Down,* a technical monograph on circadian rhythms.



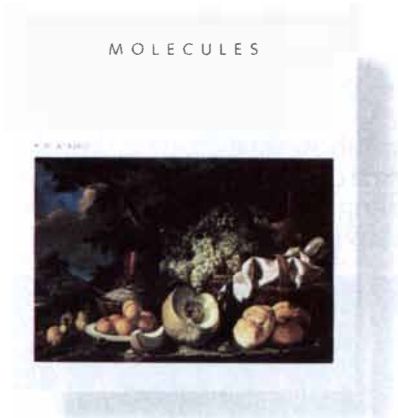
Winfree's innovative use of gradient color to express the passage of time helps us visualize the biological cycles that govern the processes of life. Courtesy of Arthur Winfree.

199 PAGES, 233 ILLUSTRATIONS

© 1987 SCIENTIFIC AMERICAN, INC

MOLECULES

P. W. Atkins



Everything we breathe, touch, eat, drink, and smell is made of molecules. They determine the taste of a barbecue and the fragrance of a rose. Yet, to many of us, molecules are an unfamiliar, abstract notion, hidden far below the surface of visible reality.

In this highly readable work, P. W. Atkins now makes the molecular world understandable, accessible, and familiar. You will meet the molecules that make sugar sweet, that make soap "soapy," and that give rise to the brilliant palette of colors unveiled in autum foliage. You will see molecules that can quell pain and alleviate depression as well as those that can induce addiction and wreak havoc on our environment.

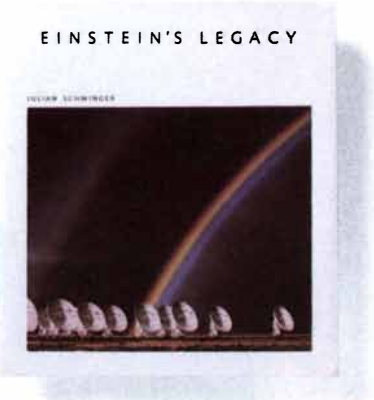
Lavishly illustrated and filled with original insights, MOLECULES portrays the structures of the substances that make up our everyday world and our daily experiences, showing us their shapes, their interrelations, why they behave as they do, and how they carry out their tasks. It gives us a more imaginative appreciation of the richness, economy, and variety of nature.

P. W. Atkins lectures in physical chemistry at the University of Oxford. He is the author of *The Second Law* (for the Scientific American Library) and the widely used textbook *Physical Chemistry,* now in its third edition.

256 PAGES, 290 ILLUSTRATIONS

EINSTEIN'S LEGACY

Julian Schwinger



Nobel laureate Julian Schwinger tells the captivating story behind one of the twentieth century's greatest achievements, the theory of relativity.

The groundwork was laid first by Isaac Newton and then by James Clerk Maxwell. But it was left to the sixteen-year-old Einstein to glimpse, for the first time, that the unifying theories of these two giants were incompatible.

In a lively and compelling narrative, Schwinger recounts Einstein's quest for the reconciliation of the conflict—a quest that ultimately led to the unification of matter and energy and of space and time in Einstein's special and general theories of relativity.

Schwinger carries this story to the space-age experiments that have been designed to confirm Einstein's legacy—and possibly lead to a still more comprehensive theory.

Julian Schwinger was awarded the Einstein Prize in 1951, the National Medal of Science in 1964, and the Nobel Prize for Physics in 1965. He is currently University Professor of the University of California.

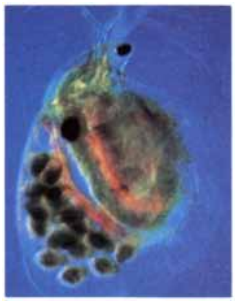
"Delightful and instructive.... An ideal gift for the curious non-expert."—NATURE

250 PAGES, 200 ILLUSTRATIONS

"The difference [between this and other book series] comes in the quality of the

"Distinguished researchers take the trouble to talk simply and directly about the work they

The largest forms of life reproduce at a slower rate and live much longer than smaller organisms. On Size and Life *examines why.*




A molecular view of the plasma membrane. Elect
axopod of the heliozoon Echinosphaerium nucle

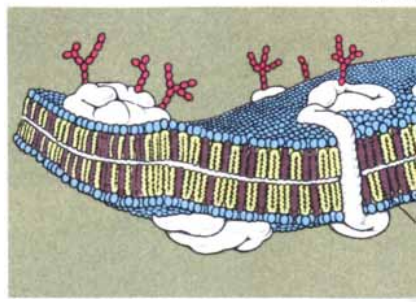
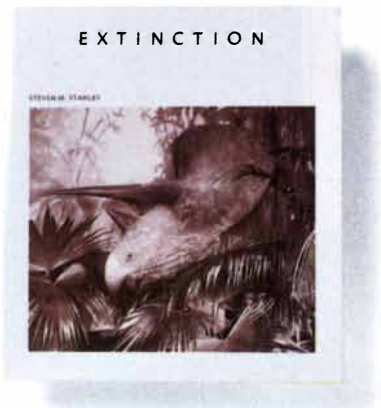


PHOTO CREDITS—Daphnia: James Bell, Photo Researchers, Inc./Hummingbird: Crawford H. Greenewalt, VIREO/Robert Wadlow, 8′ 9 1/2″ tall, and his father: Wide World Photos/Elephant: Peter J
Black Star/Important anatomical divisions of the brain: Painting by Bill Andrews/Coca shrub, source of cocaine: W. H. Hodge, Peter Arnold, Inc.

© 1987 SCIENTIFIC AMERICAN, INC

EXTINCTION

Steven M. Stanley



Since the rise of multicellular life, a handful of mysterious cataclysms has swept our planet. These geologically brief outbreaks of mass extinction have decimated tens of thousands of thriving species, from huge dinosaurs to microscopic algae.

In EXTINCTION, Steven M. Stanley gives us a fascinating account of these dramatic global disasters, investigating what the rich storehouse of rock and fossil clues—and the revolutionary findings of plate tectonics—can tell us about their origins, progress, and implications.

In light of recent speculation that meteor impacts may have ended the reign of the dinosaurs through an early version of "nuclear winter"—about which Stanley has much to say—EXTINCTION is a timely as well as intriguing geological detective story.

Steven M. Stanley, professor of paleobiology and director of graduate studies at Johns Hopkins University, is a Guggenheim Fellow and winner of the Schuchert Award of the Paleontological Society.

242 PAGES, 183 ILLUSTRATIONS

OTHER VOLUMES IN THE LIBRARY:

MATHEMATICS AND OPTIMAL FORM
Stefan Hildebrandt and Anthony J. Tromba

ISLANDS
H. W. Menard

PERCEPTION
Irvin Rock

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

THE SECOND LAW
P. W. Atkins

CONSTRUCTING THE UNIVERSE
David Layzer

THE LIVING CELL
Christian de Duve

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

FIRE
John W. Lyons

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

SUN AND EARTH
Herbert Friedman

EYE, BRAIN, AND VISION
David H. Hubel (in preparation)

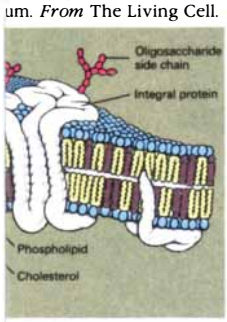
THE SCIENCE OF STRENGTH
J. E. Gordon (in preparation)

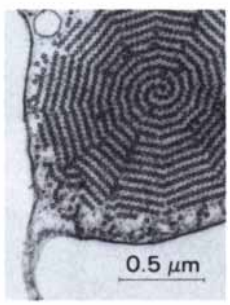
SAND
Raymond Siever (in preparation)

uthors, who combine eminence in their fields with proven gifts of exposition."
—Nigel Calder, NATURE

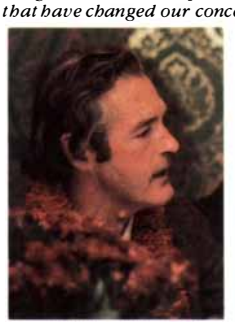
e doing and how it is solving yet another part of the mystery of the human and natural worlds."
—Steven Rose, NEW SCIENTIST

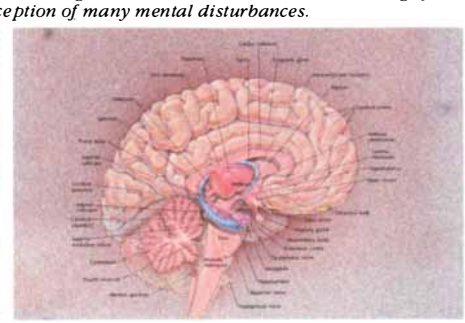
micrograph of a section through an
um. From The Living Cell.





Drugs and the Brain *explores the drugs that have led to a new understanding of the brain and that have changed our conception of many mental disturbances.*





ison, Natural History Photographic Agency/Plasma membrane illustration by Neil O. Hardy/Electron micrograph courtesy of Dr. L. E. Roth and Dr. Y Shigenaka/Timothy Leary: Gene Anthony,

© 1987 SCIENTIFIC AMERICAN, INC

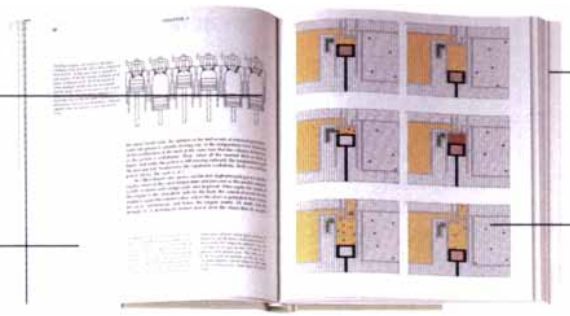
JOIN OUR LIBRARY. THERE IS NO RISK INVOLVED.

Created out of the collaboration of scientists and editors that distinguishes SCIENTIFIC AMERICAN itself, each volume in the SCIENTIFIC AMERICAN LIBRARY:

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

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

Ample margins for visual relief and notes.



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

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

YOUR PRIVILEGES AS A MEMBER

A Two-Month Cycle.
As a library member you will receive a new volume approximately every two months. Your first volume, THE TIMING OF BIOLOGICAL CLOCKS, will be shipped UPS for fastest delivery.

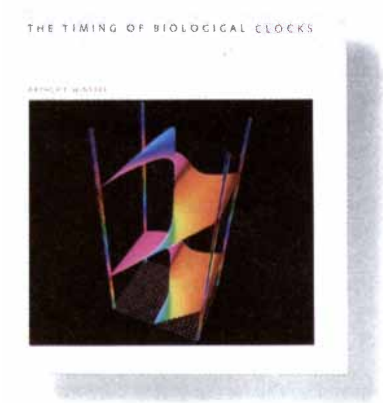
15-Day Free Examination.
Examine each new volume at your leisure. If you decide within 15 days not to own a particular volume, simply return it, and your account will be promptly credited.

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

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

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

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



MEMBERSHIP RESERVATION

Please speed my premier volume—THE TIMING OF BIOLOGICAL CLOCKS—via UPS and enroll me as a member of the Scientific American Library. Thereafter, send me a volume approximately every two months for 15 days' free examination.

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

Signature_____

Name_____

Address_____

City_____ State_____ Zip_____

0987

Send to: SCIENTIFIC AMERICAN LIBRARY, P.O. BOX 646 HOLMES, PA 19043

To join right away call toll-free 1-800-345-8112.

SCIENTIFIC AMERICAN LIBRARY

AN EXCITING NEW GENRE IN THE LITERATURE OF SCIENCE

© 1987 SCIENTIFIC AMERICAN, INC

SCIENTIFIC AMERICAN
In Other Languages

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

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

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

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

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

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

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

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

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

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

did not penetrate the intact bacterium. The solution was to link the analogue to a pair of amino acids. The amino acids are taken up by transport systems that bring needed substances into the bacterial cell, and the drug rides with them. The strategy is effective; the combination of NHdKDO and amino acids is a "potent bactericide" for gram-negative organisms, the investigators write. They add that since KDO is not present in mammalian cells, the new drug should be toxic for bacteria but quite safe for human tissues.

Although the work in Sweden is only a first step, the designer-drug approach may have broad implications. In a commentary in the same issue of *Nature* Chris Higgins of the University of Dundee argues that because the transport systems exploited by the new drug are ubiquitous, there are "endless possibilities for designer herbicides and chemotherapeutic drugs." What is more, the amino acid pair to which NHdKDO is linked is carried into the bacterium by two different transport systems. Hence two mutations would be required to block the compound's entry, which lowers the chance that resistance to the tailor-made drug will soon be acquired. *—John Benditt*

Radical Therapy

Drugs may combat brain damage by toxic oxygen free radicals

A sudden blow to the head—in an automobile accident, a fight or a fall—can lead to brain damage, but the damage does not result from the impact alone. Rather, much of the cell death probably occurs when harmful molecules loosed by the trauma launch a chemical attack on nerve cells and surrounding blood vessels. Although no effective treatment for such damage exists, several compounds are showing promise and could benefit victims of stroke and cardiac arrest as well.

The drugs work by thwarting the effects of destructive, charged molecules of oxygen called free radicals. Small amounts of oxygen free radicals are generated by ordinary cell metabolism; a severe head injury disrupts normal metabolic processes, unleashing a torrent of radicals. Similarly, brain cells recovering from a stroke or cardiac arrest also produce an excess of free radicals. In large numbers the free radicals cause a cascade of reactions that disrupt DNA, cripple enzymes, sabotage molecular gates and weaken and punch holes in cell membranes.

Investigators have reasoned that brain damage might be reduced if these secondary effects could be minimized. A number of years ago a search began for compounds that could scavenge, or neutralize, oxygen radicals. Of the many candidates, only a few seem suitable for use in the brain. Chemists at the Upjohn Company confronted the problem by designing a class of novel compounds they call lazaroids. Preliminary studies published in the *Journal of Biological Chemistry* suggest that, in addition to their scavenging activity, lazaroids penetrate cell membranes to protect them from free-radical attack. Upjohn hopes to test the drugs in clinical trials this fall.

There is also renewed interest in probucol, a substance that has been prescribed for more than 10 years to lower cholesterol levels. Recent investigations of the drug's mechanism have focused on its role as a powerful oxygen-radical scavenger. The manufacturer of Probucol, Merrell Dow Pharmaceuticals, Inc., has embarked on a program intended to gauge its potential in a spectrum of clinical applications.

One of the properties that recommend both lazaroids and probucol is their ability to cross the blood-brain barrier. Workers at the Johns Hopkins medical institutions have described encouraging results with a molecule too large to pass through the barrier. The substance, an enzyme called superoxide dismutase (SOD), is an old favorite with oxygen-radical investigators; although new to brain research, it is known to reduce free-radical damage dramatically in many other tissues. Gregory B. Bulkley, Reginald J. Davis and Richard J. Traystman reported at a recent meeting that cats deprived of oxygen for half an hour regained significant brain responses when they were treated with SOD.

Although these results spark cautious optimism, all such experiments have inevitable pitfalls. Experimental models are far removed from clinical reality: Upjohn investigators, for instance, measure brain damage in mice by the animals' ability to grasp a string. Furthermore, data implicating oxygen-radical damage in stroke and severe head injuries are plentiful but almost entirely circumstantial. Hence tests of drug efficacy are indirect at best. *—Karen Wright*

© 1987 SCIENTIFIC AMERICAN, INC

Electric Eyes

Microelectronic devices give astronomers new powers of seeing

Microelectronics is now transforming the tools of optical astronomy on the largest and smallest scales. As computer control makes it possible to build and operate ever larger telescopes, light-sensitive semiconductors will ensure that fewer of the photons such instruments can gather in unprecedented quantity go to waste. Thus equipped, astronomers will be able to glimpse objects so distant that the faint light reaching the earth left them when the universe was less than half its current age. They will also be able to take optical spectra of stars in nearby galaxies, peer through the clouds of gas and dust that veil galactic centers and perhaps even see planets circling nearby stars.

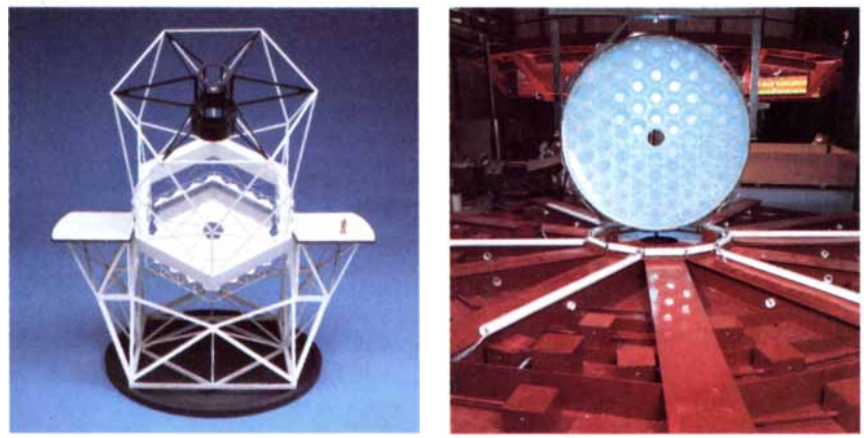
Over the past decade semiconductor light detectors called charge-coupled devices, or CCD's, have replaced photographic film as the standard medium for recording images of astronomical objects. Much more sensitive than the fastest films, CCD's detect as many as 70 percent of the photons that strike them by accumulating electrons produced by the collisions. At present the devices respond to only a limited range of wavelengths, but fluorescent coatings promise to extend the wavelength sensitivity by converting incoming radiation into a frequency CCD's can detect. Increases in the number of picture elements per device will expand their field of view; many astronomers are anxiously awaiting a new CCD with more than four million picture elements, made by Tektronix, Inc. Its appearance has been delayed by mysterious development problems.

New technology may also speed the analysis of light from distant objects. Collecting enough light for spectroscopy can require hours of observing, but new systems, by automating parts of the process, can take spectra from many objects at the same time. In one arrangement the light from 50 or more objects in a telescope's field of view is channeled along optical fibers to a spectrograph. Newer fiber-optic systems, such as one developed by John Hill at the Lick Observatory, rely on a robot to move the fibers into their correct positions in the focal plane of the telescope.

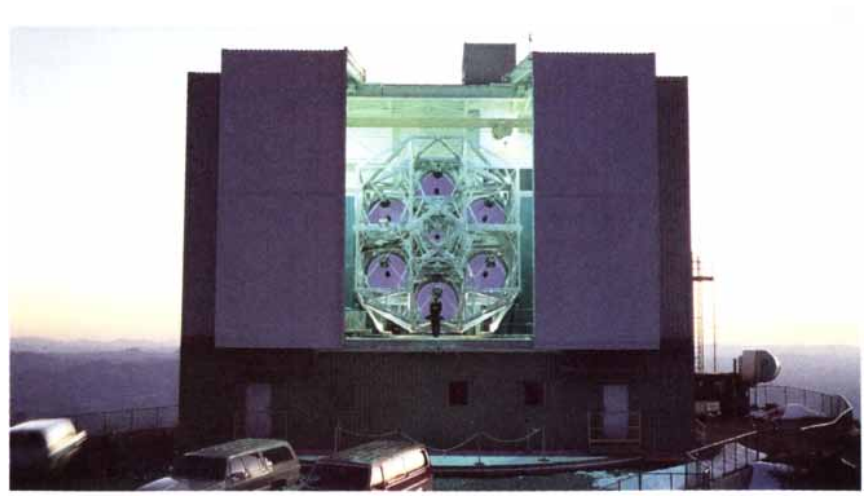
Meanwhile the telescopes themselves are becoming more powerful. A telescope's light-gathering ability can be improved by enlarging its primary mirror, but there must be no compromise in optical quality: the shape of the mirror surface must be accurate to a few millionths of an inch. The world's largest optical telescope, a six-meter instrument in the Caucasus Mountains of the U.S.S.R., is said to produce mediocre images because of an imperfect mirror.

Achieving the needed tolerances over a surface eight meters in diameter—the mirror width of several planned telescopes—requires techniques that have become feasible only with the advent of fast computing. A single mirror of that size would sag unless it were very thick, which would entail large and expensive telescope bearings to cope with the weight. The W. M. Keck telescope, now being built by the University of California and the California Institute of Technology on Mauna Kea in Hawaii, will get around the problem by using 36 1.8-meter hexagonal mirrors to form the equivalent of a single 10-meter mirror. Computer-controlled "active optics" will enable the mirrors to work in concert by adjusting the mirror supports in increments as small as a millionth of an inch to compensate for changing

Rigid cyclopean telescopes are giving way to instruments with many flexible eyes



W. M. KECK TELESCOPE, shown in a model (left), *will have a mirror composed of hexagonal segments. A rotating oven at the University of Arizona supports a "spin-cast" mirror* (right).



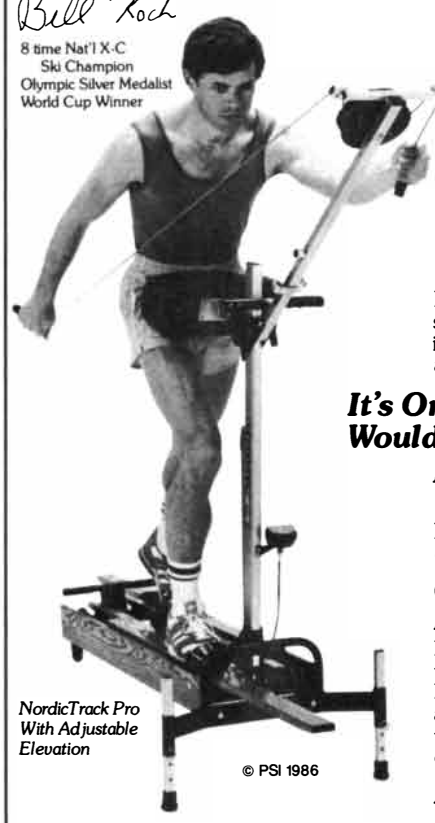
MULTIMIRROR TELESCOPE, operated by the Smithsonian Institution and the University of Arizona, is the first large telescope to use multiple primary mirrors and active optics.

© 1987 SCIENTIFIC AMERICAN, INC

Scientific Study Ranks NordicTrack #1

Bill Koch

8 time Nat'l X-C Ski Champion
Olympic Silver Medalist
World Cup Winner



NordicTrack Pro With Adjustable Elevation

© PSI 1986

In tests of exercise efficiency at a major university, NordicTrack burned more calories and provided greater aerobic workouts.

Rank of Efficiency

1. NordicTrack X-C Ski Exerciser
2. An Exercise Bike
3. A Rowing Machine
4. A Shuffle-Type Ski Exerciser

Scientific Abstract available upon request.

NordicTrack's higher oxygen usage test scores show that more muscle mass is involved in the exercise and more calories are being burned.

It's Only Logical that NordicTrack Would Get the Highest Scores.

NordicTrack is More Complete Than an Exercise Bike

NordicTrack adds important upper body exercise and more uniform leg muscle usage.

Cross Country Skiing Is More Efficient Than Rowing

NordicTrack's standing position exercises the lower body muscles more uniformly. NordicTrack's arm resistance is separately adjustable—allows a more ideally proportioned exercise. Rowing machines use only one resistance setting for both arms and legs.

Burns Up To 600 Calories per 20 Minute Workout

For effective weight control.

NordicTrack has Inertial Resistance Not Available on a Shuffle-Type Ski Exerciser.

Other ski exercisers use less effective frictional resistance with a resulting calisthenics-like feeling. NordicTrack's unique flywheel provides more effective inertial resistance for a life-like feeling of smoothness and continuity matched only by real skiing.

Free Brochure, and New Video!

800-328-5888

Minnesota 612-448-6987

NordicTrack

141F Jonathan Blvd. N., Chaska, MN 55318

WE'RE FIGHTING FOR YOUR LIFE

American Heart Association

gravitational stresses as the telescope moves.

Together the mirrors must define a paraboloidal surface; during the grinding each glass blank will be stressed so that the ground spherical surface springs into just the right shape when the stress is released. The progress has not been smooth, however: cutting blanks into hexagons causes distortions, and testing has fallen behind schedule. Project scientist Jerry E. Nelson remains confident that corrective polishing will solve the problem, and he expects the Keck telescope to see "first light" on schedule in 1990.

J. Roger P. Angel of the University of Arizona is developing a different strategy for mirror making, one that exploits the fortunate coincidence that the surface of a spinning liquid assumes the form of a paraboloid. He spin casts lightweight mirrors with a honeycomb internal structure in a revolving furnace. The technique should reduce grinding time to a few weeks, from the years required to grind a mirror from a flat blank. Because curvature varies across a paraboloidal surface, Angel is also trying to perfect a computer-controlled "lap," or grinding tool, that changes shape as it works its way across the surface of the mirror.

Spin-cast mirrors can be deeper than mirrors that are made by conventional methods; by reducing focal length this should allow telescopes to be shorter and lighter—and hence cheaper. Angel has successfully cast a 1.8-meter mirror and hopes eventually to scale up his technique to make the eight-meter mirror for the Magellan telescope, which several institutions, among them the Carnegie Institution of Washington, plan to construct in Chile.

Spin-cast primary mirrors would also be incorporated in the Columbus project, which is to be built in Arizona by a university consortium, and in the National New Technology Telescope, a gleam in the eye of Jacques M. Beckers of the National Optical Astronomy Observatories but still officially only a technology-development project. The NNTT would have four eight-meter mirrors arranged in a square on a single mount; the light from all four could be brought to a single focus, giving the telescope the light-gathering ability of a single 16-meter instrument and the resolution of a 12-meter one.

The European Southern Observatory has taken a different approach with its proposed Very Large Tele-

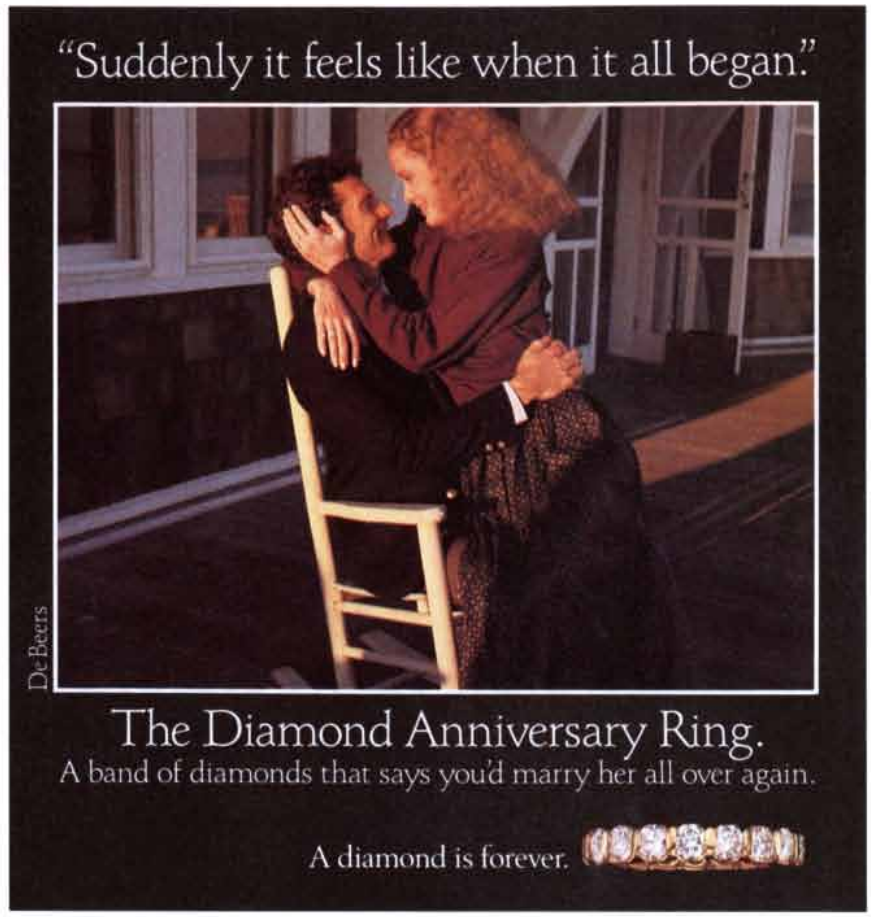
© 1987 SCIENTIFIC AMERICAN, INC

scope (which would be the world's biggest optical telescope). The telescope, to be built in Chile if it is approved, would actually consist of four eight-meter telescopes arranged in a line. They would be linked by optical interferometry to give the light-gathering ability of a 16-meter telescope, at least for observations in the infrared, where tolerances are more forgiving. This design will maximize resolving power, but combining the beams exacts penalties. Alternatively, each telescope could be used individually. The thin primary mirrors will each be supported by about 150 electronic actuators, which will adjust the shape of the mirror several times per second to compensate for wind pressure and gravitational stresses.

Future telescopes might use adaptive optics, which relies on actuators that compensate for image distortions introduced by atmospheric turbulence, the major limitation of ground-based astronomy. The computational and engineering difficulties are enormous, since adjustments must be made several hundred times per second; moreover, a large telescope might require 5,000 actuators.

Beckers is nonetheless convinced that adaptive optics will gain ground, at least for infrared work. He is testing a system that splits the image of a bright "guide star" into 37 subimages and uses their relative motions to calculate adjustments to a "floppy mirror." Stars bright enough to serve as guides are rare in the night sky, but Laird A. Thompson of the University of Hawaii at Manoa recently demonstrated that a natural layer of sodium ions in the atmosphere will scatter enough light from a powerful laser beam to create an artificial guide star.

At least one possible telescope design eschews modernity altogether. Ermanno F. Borra of Laval University in Canada has revived the old idea of a liquid telescope mirror. A thin layer of mercury would be spun on a rotating form whose shape approximates a paraboloid. In spite of the obvious disadvantage—such liquid mirrors can only look straight up—Borra claims many advantages for them, not the least of which is that liquid mirrors might be very cheap. He has tested a one-meter mirror with satisfactory results and expects to receive funds for a 2.5-meter version. Borra believes such mirrors could ultimately dwarf conventional ones: he envisions a 30-meter liquid mirror, with its form perhaps supported by rotating water. *—Tim Beardsley*



© 1987 SCIENTIFIC AMERICAN, INC

NOW DINERS CLUB HOLDS THE KEY TO BONUS FREQUENT STAYER POINTS AT HILTON, INTER·CONTINENTAL, RADISSON, RAMADA AND SHERATON HOTELS.

From now on, when you use the Diners Club Card, you can earn Club Rewards℠ points good towards exciting gifts and services. Including frequent stayer credit at the participating hotel of your choice. Which can earn you a free room upgrade or even a free suite. Or, you can choose frequent flyer miles on American, Continental, Northwest, TWA or United* airlines.

So dine with the Diners Club Card. Sleep on it. Rent with it. Whatever. And watch your frequent stayer credit grow. Or use someone else's card...and miss out on a very suite deal.

Call 1-800-DINERS-1. Join the Club. The rewards are endless.

© 1987, Citicorp Diners Club Inc.

Frequent Traveler benefits available through the Club Rewards℠ program. Participating companies subject to change. Club Rewards is a service available only to Personal Cardmembers of Diners Club from Citicorp located within the 50 United States. Free enrollment required. Certain terms and conditions apply. Free suites available at selected hotels. *United participation limited to 1986/87 Club Rewards brochure.

© 1987 SCIENTIFIC AMERICAN, INC

SCIENTIFIC AMERICAN September 1987 Volume 257 Number 3

Strategic Defense and Directed-Energy Weapons

The cochairmen of the American Physical Society panel summarize the group's findings concerning the developmental gap that stands between the laboratory and a decision to deploy such weapons

by C. Kumar N. Patel and Nicolaas Bloembergen

In November, 1983, about eight months after President Reagan called on the U.S. scientific community to develop a system that "could intercept and destroy strategic ballistic missiles before they reach our soil," the American Physical Society commissioned a study to evaluate the status of directed-energy weapons. The evaluation, which was finally released this spring, focuses on the potential of lasers and particle beams in strategic defense. The 17-member committee, of which we were the cochairmen, sought to provide a report that would serve as a technical reference point for further discussions on the feasibility of the "space shield" envisioned by proponents of the Strategic Defense Initiative (SDI), the program that grew out of the president's entreaty. This article is based on our report.

Members of the panel were selected on the basis of their expertise in the various scientific and technological fields essential for directed-energy weapons. The members were drawn from a wide range of academic, government and industrial laboratories, many of which are actively involved in the development of nuclear and related weaponry as well as supporting technologies. In preparing the study we had access to classified information. Although the public release of the report was delayed for seven months while the U.S. Department of Defense reviewed it, the amount of material deleted was minuscule. The most significant omissions in the report concern the vulnerability of the defensive systems and possible countermeasures.

What did we find? Compared with the length of the report (424 pages), our conclusions are short. We quote from the study itself: "Although substantial progress has been made in many technologies of DEW [directed-energy weapons] over the last two decades, the Study Group finds significant gaps in the scientific

Study Group of the American Physical Society report "Science and Technology of Directed Energy Weapons"

Nicolaas Bloembergen, *cochair*
Harvard University
C. Kumar N. Patel, *cochair*
AT&T Bell Laboratories
Petras Avizonis
Air Force Weapons Laboratory
Robert Clem
Sandia National Laboratories
Abraham Hertzberg
University of Washington
Thomas Johnson
U.S. Military Academy
Thomas Marshall
Columbia University
Bruce Miller
Sandia National Laboratories
Walter Morrow
M.I.T. Lincoln Laboratories
Edwin Salpeter
Cornell University
Andrew Sessler
Lawrence Berkeley Laboratory
Jeremiah Sullivan
University of Illinois
James Wyant
University of Arizona
Amnon Yariv
California Institute of Technology
Richard Zare
Stanford University

L. Charles Hebel, executive officer
Xerox Corporation
Alex Glass, principal consultant
KMS Fusion, Inc.

The history of the study is a long one. Commissioned on November 20, 1983, by the APS, the study group was assembled by November, 1984. The report required some 21 months ("the gestation period of an elephant," according to Patel) until its release on April 24, 1987. The last seven months were taken up with security reviews by the Department of Defense.

© 1987 SCIENTIFIC AMERICAN, INC

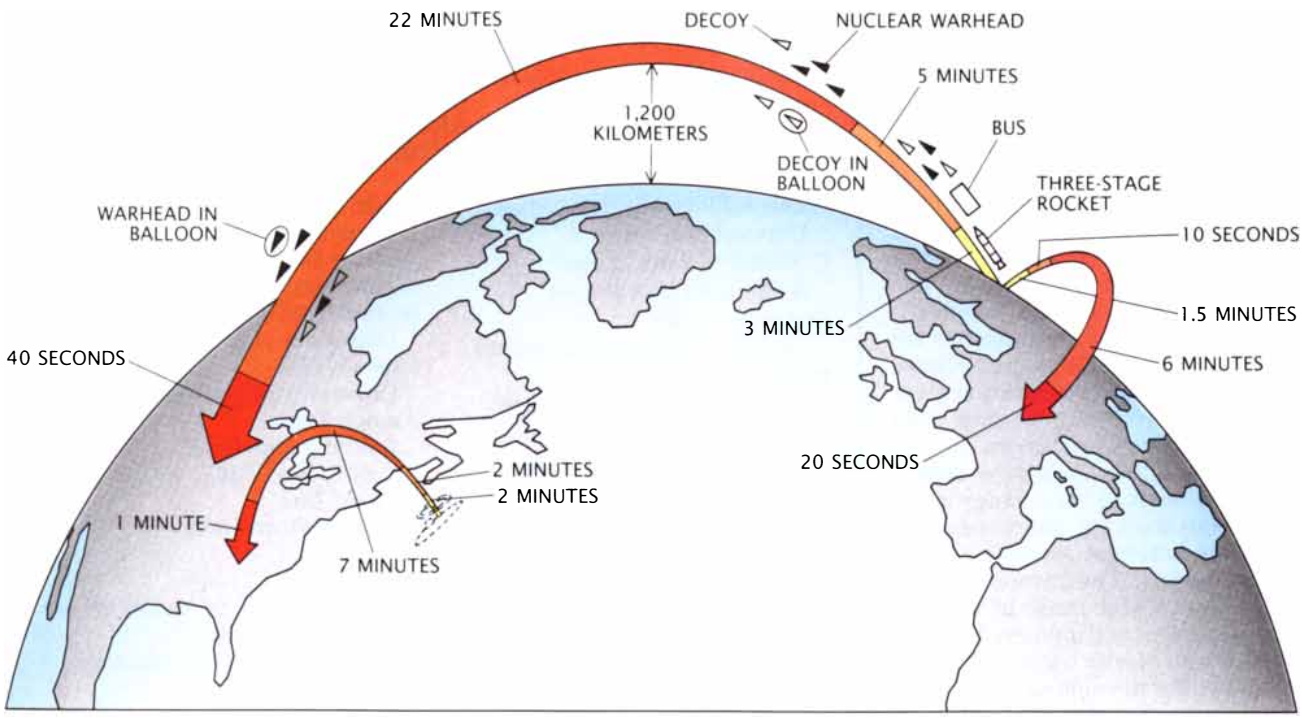
and engineering understanding of many issues associated with the development of these technologies. Successful resolution of these issues is critical for the extrapolation to performance levels that would be required in an effective ballistic missile defense system. At present, there is insufficient information to decide whether the required extrapolations can or cannot be achieved. Most crucial elements required for a DEW system need improvements of several orders of magnitude. Because the elements are inter-related, the improvements must be achieved in a mutually consistent manner. We estimate that even in the best of circumstances, a decade or more of intensive research would be required to provide the technical knowledge needed for an informed decision about the potential effectiveness and survivability of directed energy weapon systems. In addition, the important issues of overall system integration and effectiveness depend critically upon information that, to our knowledge, does not yet exist."

The study focused on directed-energy weapons because they would be needed in almost all stages of the destruction of a missile, including detecting the launch, locating and tracking the target, distinguishing warheads from decoys, destroying the target itself and verifying the kill. The study did not explicitly review the other major class of weapons, kinetic-kill weapons: chemical guns (rockets) and electromagnetic guns. A ballistic-missile defense that relied on kinetic-kill weapons for the actual destruction of a target would still need directed-energy technology to carry out the other tasks outlined above. As such, an effective ballistic-missile defense is very dependent on the availability of mature directed-energy technology.

The study also made no attempt to discuss in detail many significant issues concerning command, control, communication and intelligence (C^3I), computing hardware, software creation and reliability for battle management and overall systems complexity. Other issues that were recognized but not addressed include manpower requirements, costs and cost-effectiveness, arms control and strategic stability, as well as international and domestic policy implications. Many of these topics have been the subject of intense debate in the years following the president's "Star Wars" address; it is somewhat surprising that the one aspect that had not been analyzed with sufficient objectivity and in sufficient detail was the technology itself.

The effectiveness of any ballistic-missile defense depends on taking careful account of a missile's four phases of flight: boost, postboost, midcourse and reentry. The boost phase begins when the missile leaves its launcher and ends when the payload separates from the lifting vehicle. The boost phase usually lasts for about three minutes. During a typical postboost phase, which has a duration of about five minutes, a "bus," or postboost vehicle, ejects a number of smaller missiles called reentry vehicles (typically called multiple independently targeted reentry vehicles, or MIRV's). Thrusters are actuated to help direct the reentry vehicles to individually designated targets. Often the boost phase is defined as the total period covering the launch and deployment of all the reentry vehicles. We have chosen to divide this period into two because of the different signatures available to the defense during the two parts.

The longest phase is the midcourse flight, in which the reentry vehicles and any decoys move along together in nearly frictionless trajectories

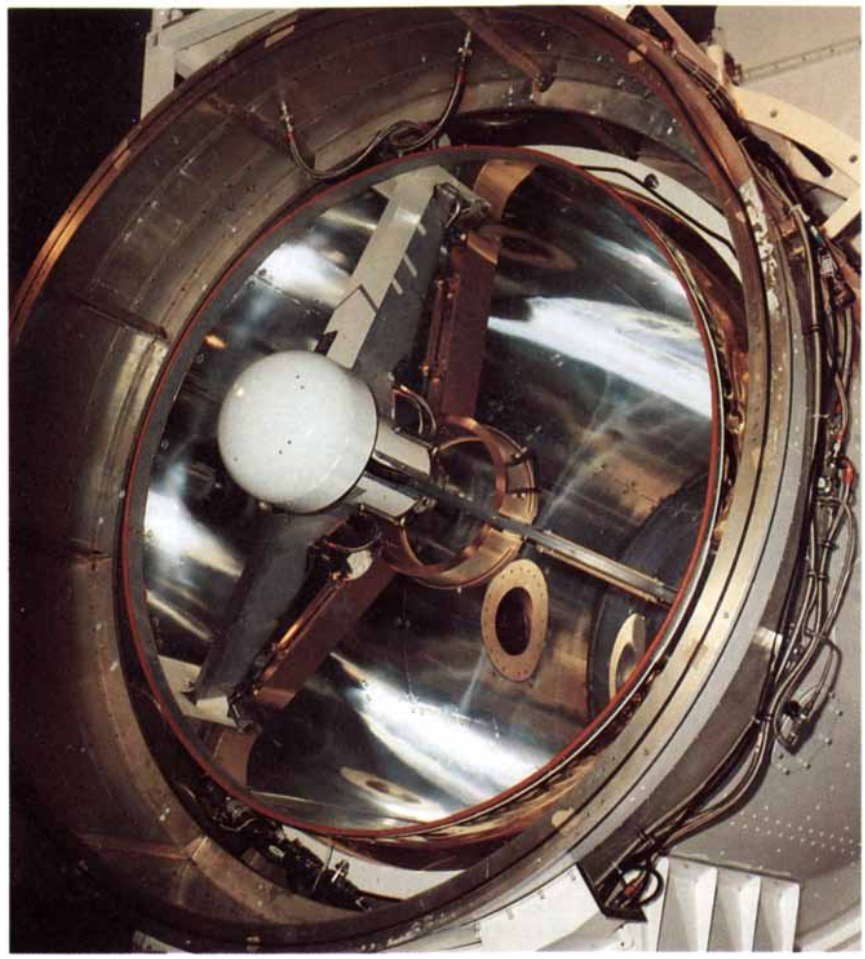


FOUR PHASES of flight are shown for an intercontinental ballistic missile (ICBM), a submarine-launched missile and an intermediate-range missile. The first phase, the boost phase (*yellow*), begins when the missile leaves the launcher and ends when the payload separates from the lifting vehicle. During a typical postboost phase (*light orange*) a "bus," or postboost vehicle, ejects a number of smaller missiles called reentry vehicles. In the midcourse phase (*dark orange*) the reentry vehicles and any decoys move along in nearby trajectories. The flight ends with the reentry phase (*red*). Because of the large number of targets associated with the midcourse phase, the best points at which to thwart an enemy attack are during the boost and postboost phases.

© 1987 SCIENTIFIC AMERICAN, INC

far above the atmosphere. The midcourse phase lasts for about 20 minutes for intercontinental flights. Finally, the reentry vehicles pass through the atmosphere; less than 60 seconds later they strike the earth.

Perhaps the best points at which to thwart an enemy attack are during the boost and postboost deployment phases. For a variety of reasons, directed-energy weapons do not have an important role in the final reentry phase. The advantage gained by the long length of the midcourse phase is offset by the increased number of threat objects (reentry vehicles and decoys) over that of the boost and postboost deployment phases. Indeed, given the present number of Soviet boosters and their capability, it is estimated that half a million or more threat objects could be deployed. Even a boost-phase defense that was 80 percent effective would still leave 100,000 or more objects entering the midcourse phase. The tracking and discrimination of tens or hundreds of thousands of objects would pose formidable challenges to sensors and battle-management computers.



LASER-BEAM DIRECTOR is designed to track objects in flight and to focus a laser beam on them. The instrument is part of the mid-infrared chemical laser (MIRACL) at the White Sands Missile Test Range in New Mexico. A chemical laser generates a beam through chemical reactions between gases. Here the gases are hydrogen and fluorine.

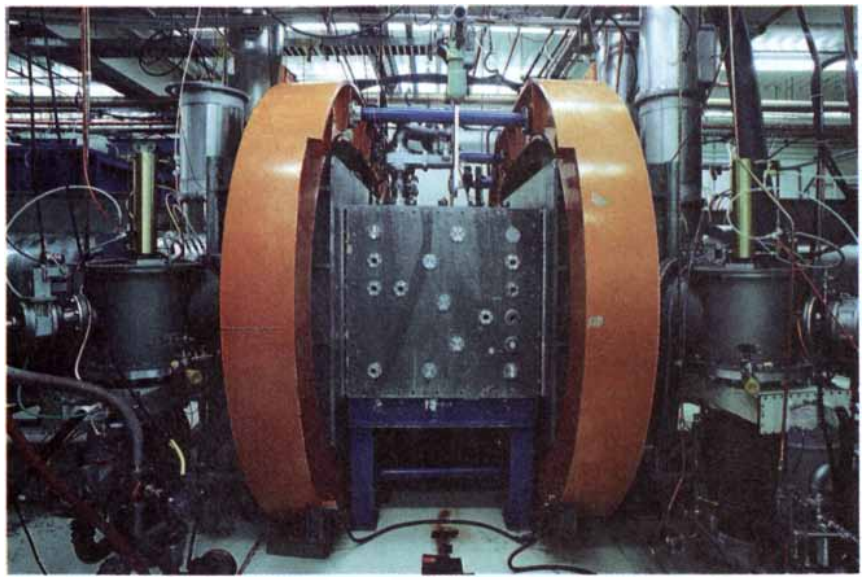
Currently four kinds of lasers are being considered as kill weapons for operation during the boost phase. They are chemical, excimer, free-electron and X-ray lasers. The beams produced by all of them travel at the speed of light, which means that for all practical purposes any target can be reached instantaneously. Chemical lasers, the maturest technology being considered, generate radiation by means of chemical reactions between two gases, such as hydrogen and fluorine, or deuterium ("heavy hydrogen") and fluorine. Running continuously, hydrogen-fluoride and deuterium-fluoride lasers have been reported to have power outputs exceeding one megawatt (10^6 watts); a 10-kilowatt laser beam can cut through a quarter-inch steel plate in a matter of seconds.

In spite of the impressive power of chemical lasers and the high quality of their beams, we estimate that the least demanding strategic-defense applications require the present power levels to be increased by at least a factor of 20 while keeping the beam free of distortion and minimizing its divergence. For a typical distance between a target and a laser the needed power may require an additional improvement by a factor of four. The chemical-laser geometry that has produced the megawatt-level power is not considered scalable to much higher powers. The needed improvement must therefore be obtained with a geometry that has not yet been explored. Whether or not a chemical laser can be made sufficiently powerful remains to be seen.

Moreover, the wavelength of light emitted by a hydrogen-fluoride laser (2.8 micrometers, or millionths of a meter) is absorbed by the atmosphere. As a result a hydrogen-fluoride laser would have to be based in space, which would lead to problems arising from vibrations and exhaust from burnt fuel. We should point out, however, that the atmosphere is virtually transparent to the beams of deuterium-fluoride and atomic-iodine chemical lasers, which have wavelengths of 3.8 and 1.3 micrometers respectively. As such the two lasers could be based on the ground if the necessary power requirements are met. It is worth noting, however, that we find oxygen lasers need to be scaled up by at least five orders of magnitude (a factor of 100,000) over their current capabilities.

Excimer lasers are the second kind of laser being considered for directed-energy weapons. Excimer stands for excited dimer: an unstable compound composed of two molecules. An electric discharge excites the molecules into forming an ionically bound dimer molecule. The dimer gives off radiation and dissociates. An excimer laser produces light in the form of short pulses. Among the most powerful excimer lasers are the krypton-fluoride lasers under development at the Los Alamos National Laboratory. Such lasers have a wavelength of .25 micrometer, which has the disadvantage of relatively poor atmospheric transmission compared with visible light or well-chosen infrared lasers. In the case of the excimer lasers of the most interest, the poor transmission is caused not by absorption but by the high amount of

© 1987 SCIENTIFIC AMERICAN, INC



EXCIMER LASER under construction at the Los Alamos National Laboratory will produce light in the form of rapid pulses. Excimer is a contraction of excited dimer, an unstable compound composed of two molecules. An electric discharge excites the molecules into forming an ionically bound dimer molecule. The dimer gives off radiation and dissociates. Here the excimer is krypton-fluoride. The large oval magnets ensure that the electric discharge (from an electron beam) is deposited evenly throughout the krypton-fluoride gas. The laser beam itself would oscillate in and out of the page.

scattering from molecules in the atmosphere. (The particular type of scattering, known as Rayleigh scattering, increases as the inverse of the fourth power of the radiation wavelength.) The problem can be ameliorated somewhat by shifting the excimer-laser wavelength to longer wavelengths through a technique called Raman shifting.

We estimate that ground-based excimer lasers for strategic-defense applications must produce at least 100 megajoules of energy in a single pulse or train of pulses with a total duration of between several microseconds and several hundred. Existing pulsed excimer lasers can generate about 10 kilojoules of energy in a pulse lasting for about a microsecond: the energy needs improvement by at least four orders of magnitude. The gap might be bridged by combining many lasers, but the feasibility of such a scheme is unclear.

The third kind of laser under development that could serve in ballistic-missile defense is the free-electron laser. It operates by sending a beam of electrons through a series of "wiggler" magnets that cause the electrons to vibrate and emit radiation. By changing the distance between the magnets or the energy of the electrons, the laser can be tuned to radiate at theoretically any wavelength. For ballistic-missile defense applications a ground-based free-electron laser should have an average power of at least one gigawatt (one billion watts) at a wavelength of one micrometer. Radiation of such a wavelength would pass through the atmosphere with ease.

Peak powers of approximately a megawatt have been produced at a wavelength of one micrometer; peak powers of approximately a gigawatt have been produced at a wavelength of eight millimeters (a wavelength absorbed by the atmosphere). Scaling to short wavelengths at high powers is a difficult technical problem. The feasibility of building high-efficiency, high-power free-electron lasers that operate at one micrometer depends on first verifying several physical concepts that have been developed only theoretically.

Perhaps the most exotic kind of laser under development is the X-ray laser. The device consists of a nuclear explosive surrounded by a cylindrical array of thin metal fibers. X rays emitted during the nuclear explosion stimulate the emission of a beam of X rays from the fibers in the short time before the device destroys itself. Workers have tested the X-ray laser in an underground site, but the feasibility of making a militarily useful X-ray laser remains uncertain. The absorption of X rays by the atmosphere means the device would have to be deployed at an altitude greater than about 80 kilometers, perhaps in some kind of "pop up" scheme. A way must be found to focus and guide the beams of X rays toward their targets. Many other physical concepts must be validated before the application of nuclear-pumped X-ray lasers to strategic defense can be evaluated.

Even if lasers sufficiently powerful for strategic-defense applications can be made, significant hurdles must be overcome to deliver the beams to their targets. First, any laser beam, no matter how intense and collimated it is, will suffer from diffraction as it travels through space: the beam will spread and become less intense. For a given power output from a laser, the intensity of the beam on a target is proportional to the square of the diameter of the mirror with which the beam is focused. The intensity is also inversely proportional to the square of the product of the wavelength of the laser and the distance to the target. Consequently the longer the wavelength or the range is, the larger the diameter of the mirror must be to maintain the desired target intensity.

The largest mirrors that are practical for steering and pointing have diameters of about eight meters, but diameters of 10 to 40 meters would be required for ballistic-missile defense missions. Such effective sizes could be achieved by employing an array of small mirrors instead of a single large mirror. The mirrors would have to be coordinated by electrically driven devices called actuators. Although it is currently possible to control several hundred actuators at once, it is not known whether the estimated 10,000 to 100,000 actuators for defense tasks can be controlled simultaneously. Moreover, the array of mirrors calls for phase-correction techniques, in which the "crests" and "troughs" of the waves in a beam are carefully aligned. It remains to be seen whether such techniques can work for the high-intensity beams needed for defense purposes. An alternative approach, which would make use of a single, large, flexible membrane, is in the conceptual stage of development.

The mirrors in any laser system would be vulnerable to radiation from other lasers, particularly if the mirrors were part of a space-based system. Even a relatively weak laser could cause significant damage if its

© 1987 SCIENTIFIC AMERICAN, INC

wavelength were different from the one the coating was designed to handle. Energetic particles from cosmic rays could also damage the coating. Small mirrors in the laser would have to be cooled to prevent damage from the laser beam itself.

Ground-based lasers, such as free-electron lasers and excimer lasers, have advantages over space-based ones in that weight, power and servicing problems are not relevant. Even ground-based lasers, however, would depend on substantial optical components mounted on space platforms for relaying the laser radiation from a ground site in the continental U.S. to a target not within the line of sight. In addition atmospheric turbulence will degrade laser-beam quality. A technique called optical phase conjugation is being explored as a way of compensating for the turbulence. In this technique one would measure the distortion of a low-power beam from a beacon laser on a relay-mirror platform in space. The information would be encoded in the outgoing high-power laser beam from the earth in such a way that the beam is "predistorted" and emerges from the atmosphere with its original high quality. The technique has been demonstrated at low powers, but it needs to be scaled up to high ones. Furthermore, the number of actuators needed to deform the mirror that would produce the predistorted beam must be scaled up by two orders of magnitude.

During the operation of a pulsed laser (such as an excimer laser) the optical field associated with the downward leg from the mission mirror to the target would be quite intense. The intensity would be high enough to cause the laser beam to scatter (through a process called stimulated Raman scattering) at altitudes below 80 kilometers. The phenomenon changes the wavelength of the radiation, which in itself is not crucial, but it also reduces the power available for attacking the target.

Finally, ground-based laser systems would have to be set up at multiple sites separated by hundreds of kilometers to keep adverse weather conditions, such as cloud cover, from immobilizing the defense. Each of these sites in turn would require some duplication of large telescope mirrors over scales of a few kilometers in order to deal with local cloud-cover conditions.

The other class of directed-energy weapons being considered for ballistic-missile defense consists of particle beams, which can be made up of electrically charged particles or of electrically neutral ones. Most charged-particle beams consist of high-energy electrons. We estimate that booster-kill applications of a charged-particle beam call for a scale-up in accelerator voltage by at least one order of magnitude, in pulse duration by at least two orders of magnitude and in average power by at least three orders of magnitude. Discrimination between targets and decoys requires a scale-up in pulse duration by at least two orders of magnitude and in average power by at least two orders of magnitude.

Charged-particle beams have the limitation that they are bent by the magnetic field of the earth. Moreover, they tend to "blow up," or be unstable, as a consequence of the repulsive force between charged particles that have the same sign. Attempts to compensate for both problems have been made using laser beams. The basic idea is to create in a gas an ionized channel that guides a high-energy electron beam, just as an optical fiber guides a light beam. Such a channel is formed by directing a laser beam through the gas to strip the electrons from the gas atoms. This mechanism for beam guiding has been successfully demonstrated in the laboratory over distances of 95 meters; it would have to be effective over distances of 1,000 kilometers for ballistic-missile defense applications. Furthermore, the laser-created plasma channel for guiding electron beams cannot be used at altitudes so high that there is little gas to ionize but where the effects of the earth's magnetic field are still felt. Also, at low altitudes the high density of gas results in beam instabilities.

Because neutral-particle beams are not affected by magnetic fields, investigators have turned toward developing them. To generate a neutral beam, a beam of negative ions (atoms that have an excess electron) is first accelerated to the required energy, directed and focused, and the extra electron is then stripped away. We

FREE-ELECTRON LASER called PALADIN at the Lawrence Livermore National Laboratory radiates in the infrared portion of the electromagnetic spectrum. A beam of electrons is accelerated to energies of 50 million electron volts (MeV). The electrons then pass through the series of "wiggler" magnets shown here. The magnets cause the electrons to vibrate and emit radiation.

© 1987 SCIENTIFIC AMERICAN, INC

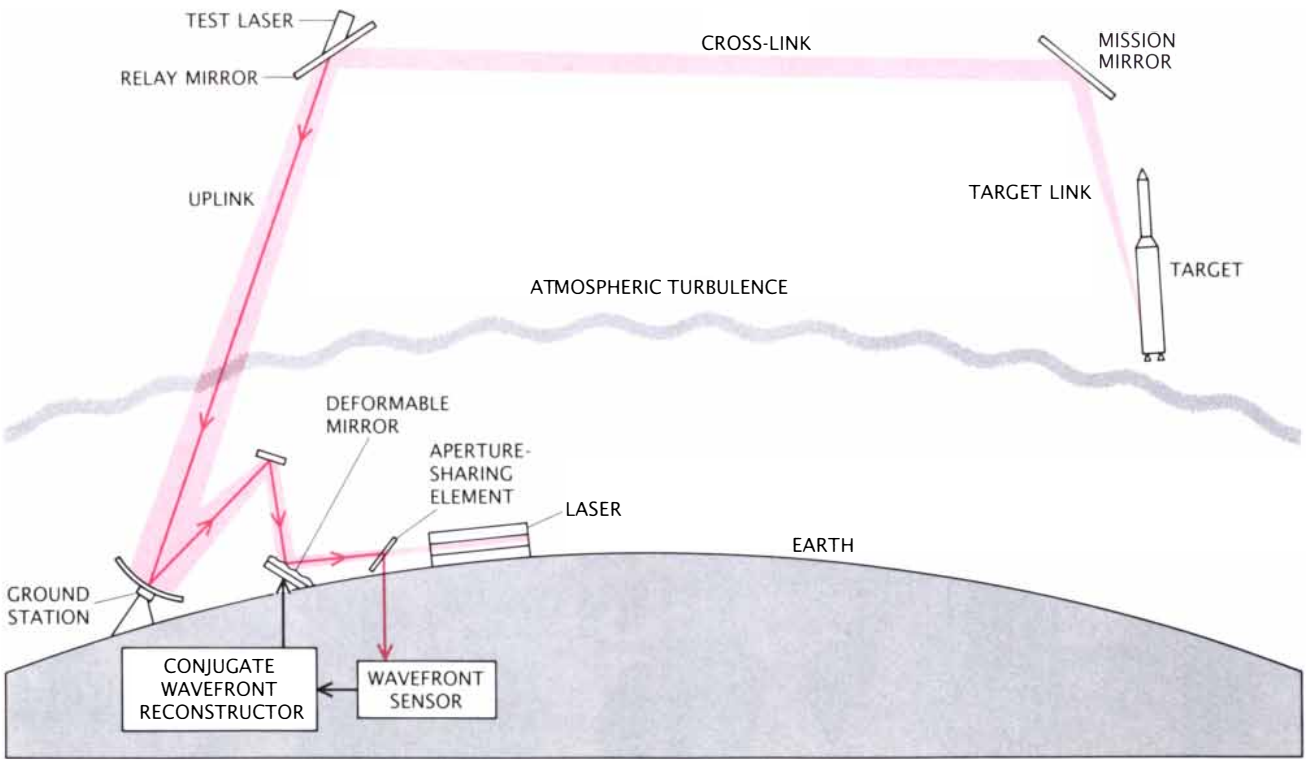
estimate that neutral-particle beam accelerators operating at the necessary current levels must be scaled up by at least two orders of magnitude in both voltage and the rate at which they can deliver a beam. The pointing accuracy and the rate at which the devices can be retargeted must be improved considerably. Another drawback of neutral-particle beams is that they interact strongly with all matter. At lower altitudes, where the gas density is substantial, the atmosphere strips the neutral particles of their outer electrons. As a result a neutral-particle beam can become a charged-particle beam and inherit the latter's limitations. Neutral-particle-beam devices would have to be based in space.

No matter how a strategic-defense system seeks to kill a missile—be it with lasers, particle beams, rockets or electromagnetic guns—that system can be no more effective than its ability to detect and track its target. In order to ensure that 90 percent of the incoming missiles are destroyed during the boost and postboost deployment phases, for instance, well over 90 percent of the missiles must first be detected. In addition the ability to track and destroy a target during the midcourse phase depends on knowing with high precision the target's trajectory during the boost phase. Of even greater importance is the need to maintain low false-alarm rates so that a defense system against ballistic missiles is not activated in peacetime.

A missile is typically tracked during its boost phase by detecting the intense infrared radiation from its booster plume. The position of the missile within the plume depends in a complex way on altitude, missile type and the kind of rocket motor and fuel. As a consequence the precise location of the missile is susceptible to variation by the offense that cannot be predicted by the defense. Infrared tracking of missile plumes will have to be supplemented by other means to ensure that the precise aiming requirements of directed-energy weapons are met.

Tracking requirements also pose a challenge during the postboost and midcourse phases. Because of the weak thermal signatures of postboost and reentry vehicles, thermal detectors will have to be supplemented with microwave or optical radars. A large number of space-based platforms carrying the detectors would be needed. Such platforms could perhaps contain supplementary detectors for tracking during the boost phase as well.

In the midcourse phase an additional challenge confronts the defense: reentry vehicles must be distinguished from decoys, and many options are available for confusing or saturating defensive detectors. Directed-energy technologies may offer the possibility of identifying decoys. Particle beams penetrate deep into all kinds of material; a neutral-hydrogen beam at from 100 to 400 million electron volts (MeV) of energy, for example, can penetrate from four to 40 centimeters of aluminum. Hence particle beams directed at a target can sample its contents. The kinds of emissions from the target's interior could be exploited to determine whether it is a decoy. Such discrimination would require large numbers of additional directed-energy platforms based in space. The



ATMOSPHERIC TURBULENCE, which reduces the quality of laser beams, could be compensated for by adaptive optics, in this case a computer-driven deformable mirror. A signal from a weak laser beam directed from the relay mirror to the ground tells the wavefront sensor the nature and extent of the atmospheric distortion. A computer then instructs electrically driven devices called actuators to deform segments in the mirror, so that the more powerful laser beam from the ground is launched with a "conjugate distortion": when the beam propagates through the atmosphere, the distortion is "undone" and the beam reaches the relay mirror undistorted. Although it is currently possible to control several hundred actuators at the same time, it is not known whether the estimated 10,000 to 100,000 actuators needed for purposes of defense can be controlled simultaneously.

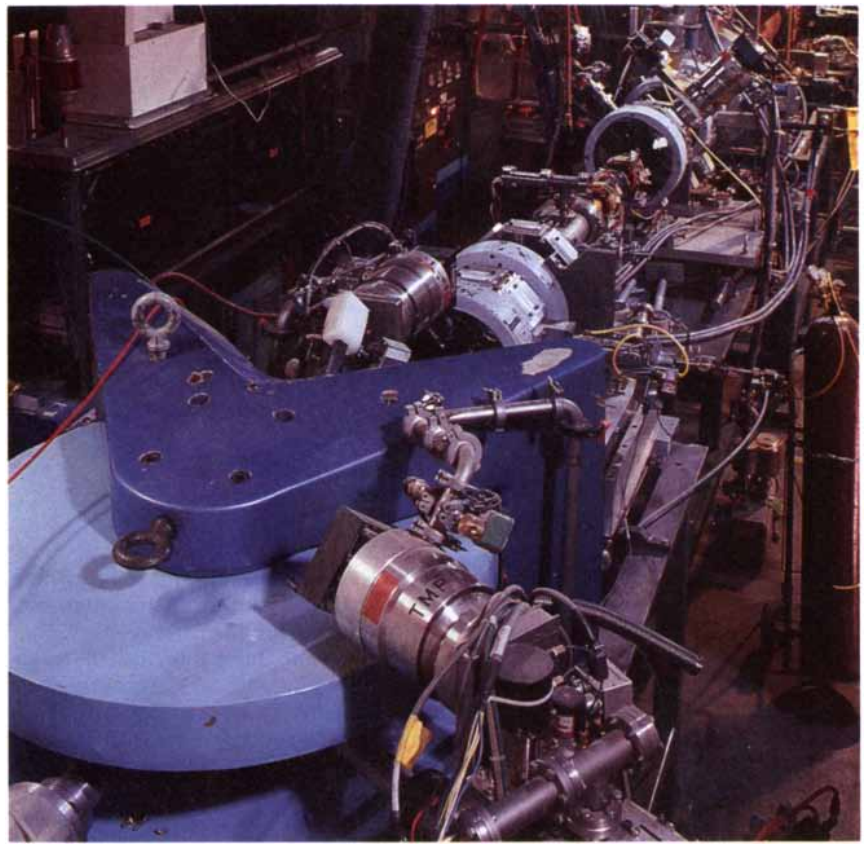
© 1987 SCIENTIFIC AMERICAN, INC

platforms would have to be able to operate in an environment that might contain large amounts of nuclear radiation. The application of directed-energy weapons to discrimination is currently in the conceptual and early experimental stages.

Any strategic-defense system utilizing directed-energy weapons would require significant amounts of power. A space platform would need from about 100 to 700 kilowatts of continuous power to satisfy "housekeeping" functions alone—to control altitude, cool mirrors, receive and transmit information and operate radars. Since no precise designs for these platforms exist, the requirements stated here should be considered reliable only to within a factor of two. In any case such a power level could be attained effectively only with a nuclear reactor. Each platform would need its own reactor, so that perhaps 100 or more reactors would have to be deployed in space. Meeting the challenge means first solving many daunting engineering problems that have not yet been explored, such as how to cool large space-based nuclear reactors.

The power requirements during an actual engagement could reach from 100 megawatts to a gigawatt for periods of several hundred seconds, depending on the type of space-based directed-energy weapon. The power would have to come from large chemical or nuclear rocket engines, which would have to be deployed at considerable distances from the platforms to avoid mechanical disturbances. If chemical engines were used, the fuel consumption would be more than five tons per minute of operation per platform. The system for transferring energy from the engines to the platforms would no doubt be complex. The prime power requirements for space-based directed-energy weapons present significant technical obstacles.

Another key issue for any ballistic-missile defense system is survivability. Space platforms would carry such delicate instruments as sensors, optical mirrors and radar dishes, many of which have considerably lower damage thresholds than boosters, postboost buses and reentry vehicles. Although sensors and optical mirrors can be shielded during long periods of inactivity, they would be exposed when put on the alert prior to an impending attack. The defense system would be vulnerable to assault by space- and ground-based directed-energy weapons and kinetic-energy weapons. The system would be particularly vulnerable to attack in the course of being assembled in space. The survivability of a defense system employing space-based assets is highly questionable.



NEUTRAL PARTICLE BEAM at Los Alamos is produced by first accelerating a charged beam of negative ions (atoms that have an additional electron). The extra electron is then stripped in a gas cell, leaving a neutral beam. The large blue chamber, which is positioned at the end of the accelerator, is used for measuring the energy of the particles.

The survivability of ground-based facilities also raises serious issues. The facilities would have to be protected successfully from direct attack by many threats, including cruise missiles and even sabotage. The projected small number of ground-based facilities, each of which would have to be capable of carrying out the entire task of the directed-energy weapon component of a ballistic-missile defense, would put a high premium on these sites.

Survivability is further called into question by the simple observation that even if a directed-energy weapon is too weak to serve in a ballistic-missile defense, it may still be powerful enough in the hands of the offense to threaten space-based components of a defensive system. Space-based platforms move in known orbits and can therefore be targeted over longer time spans than boosters, postboost buses and reentry vehicles. The platforms may have key components that are more vulnerable than boosters and reentry vehicles. Space-based platforms in low orbits can also be attacked from shorter ranges than those required for boost-phase intercepts. Moreover, X-ray lasers driven by nuclear explosions would constitute a special threat to the delicate operation of space-based sensors, electronics and optics.

The issue of strategic and tactical environment should also be considered. The development and deployment of an effective ballistic-missile defense would occupy many years. As a result considerable time will be available during which the offense can develop countermeasures. Any defense will have to be designed to handle a variety of such responses, yet it seems possible that some unanticipated ones could be deployed. It is quite conceivable that a directed-energy weapon system designed for today's threats will be inadequate for one or more of the threats it will face when it is deployed.

© 1987 SCIENTIFIC AMERICAN, INC

The Large-Scale Streaming of Galaxies

The Milky Way is traveling through the universe in concert with a swarm of other galaxies. The source of the impetus may be a remote concentration of mass on a scale that challenges current theory

by Alan Dressler

The origin and fate of the universe are imprinted on the motions of its constituents. In the early 1920's, soon after the discovery that our own galaxy is one of billions, it was shown that distant galaxies are receding in all directions. It was evident that the universe is expanding and the galaxies are being swept apart, carried by a uniformly enlarging fabric of spacetime.

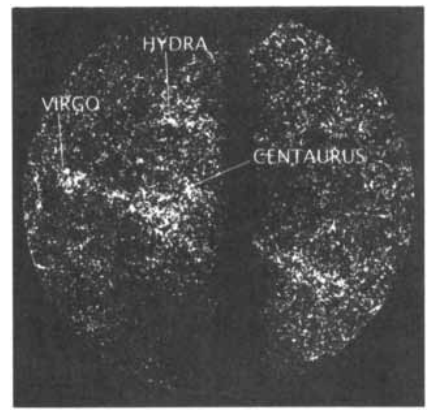
More recently it has become clear that galaxies are not at rest with respect to this unfolding fabric. They show motions of their own, which offer insight into the texture of the cosmos: the extent of its fluctuations in mass density. What is already known of the motions suggests that matter is clumped together on unimaginably large scales, reflecting poorly understood events in the early universe. The motions may also answer a question about the other end of time: Will the expansion of the universe continue, or will the force of gravity ultimately stop the expansion or even reverse it, causing the universe to recollapse into primordial density? Several groups, including a group to which I belong, are engaged in mapping the pattern and scale of the cosmic flow.

The cosmological importance of galactic motions was established in the early part of this century, when the American astronomers Edwin P. Hubble and V. M. Slipher began measuring displacements of spectral lines in the light from other galaxies. Such shifts reflect the Doppler effect: the change in the wavelength of radiation or sound that results from motion of the source with respect to the observer. Most of the galaxies showed a red shift, meaning that the lines were displaced toward the red, or longer-wavelength, side of the spectrum. A red shift indicates motion away from the observer.

Sixty years later it is difficult to appreciate how radical suggestions that other galaxies are receding in all directions sounded when they were made in the early 1920's. They represented a change in world view in every way as fundamental as the Copernican revolution; the notion of a static universe was by then as well established as the geocentric universe had been in the 15th century. Indeed, Hubble himself at first interpreted his data as evidence for a motionless universe. He believed he had confirmed a theoretical analysis by the Dutch mathematician Willem de Sitter, who had found a static solution to Einstein's equations of general relativity. By including a term that described a dilation of time with increasing distance, the analysis gave rise to red shifts that increased for more distant galaxies, even without relative motion.

By the late 1930's, however, the idea that the universe is expanding had seized nearly everyone, even the recalcitrant Hubble, and measuring the red shifts of galaxies at ever greater distances was the order of the day. When the red shifts of remote galaxies were correlated with their distances, estimated from their apparent brightnesses, a relation emerged that was as simple as one could have hoped: the recession velocity of a galaxy equals its distance multiplied by a constant. The implication of this linear relation, known as the Hubble law, was clear: the expansion of the universe is uniform. A being in a galaxy billions of light-years away would see the same relation between distance and red shift in surrounding galaxies as is seen from the Milky Way. By the 1970's it was generally accepted that the Hubble law has a second implication: the observed expansion of the universe was unleashed some 15 billion years ago in an explosive creation event, the big bang.

Even as red-shift data for distant galaxies improved, the correlation between distance and red shift remained disappointingly poor for nearby galaxies. Most workers believed the poor correlation reflected



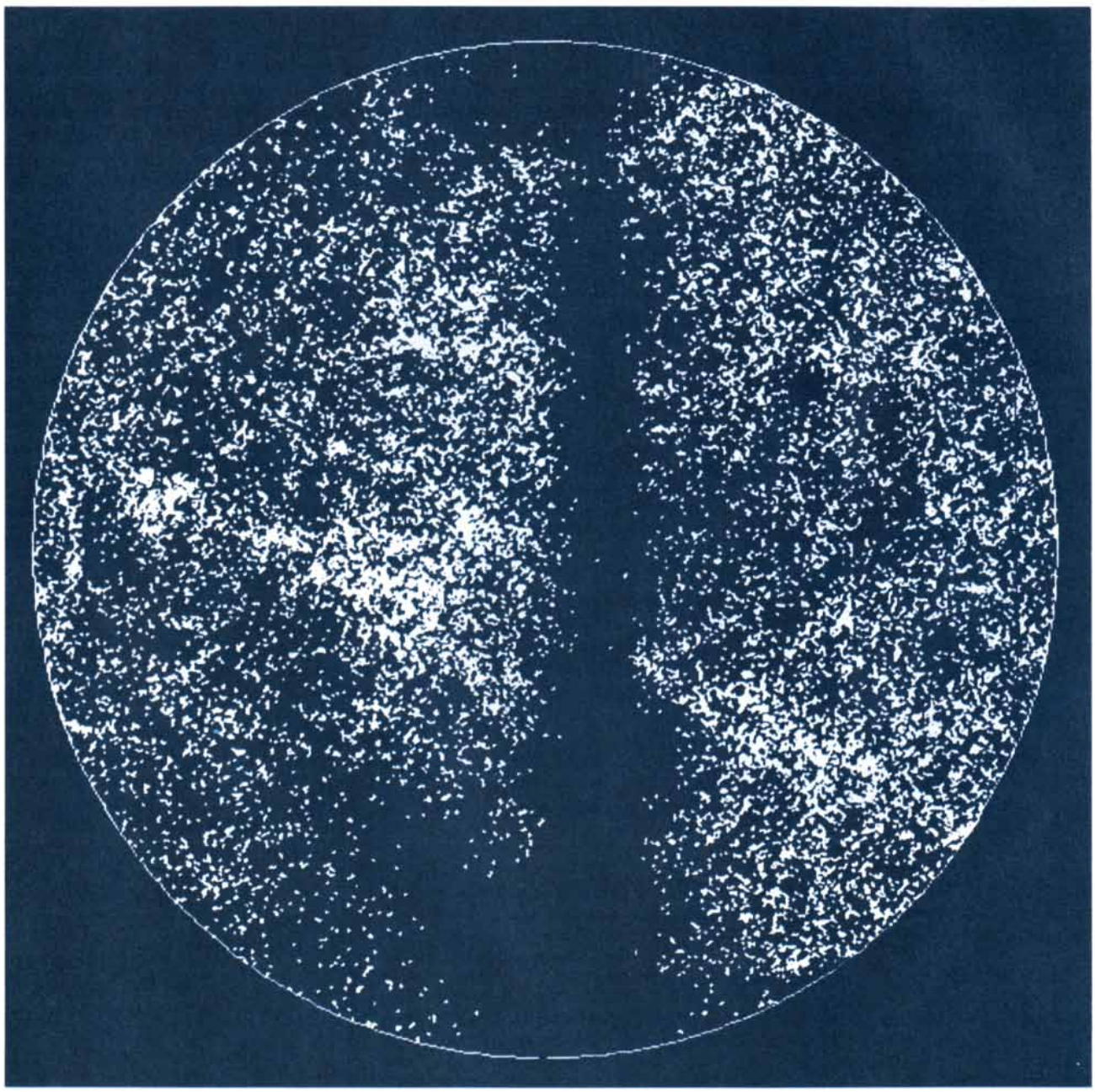
CELESTIAL HEMISPHERE showing the distribution of observed galaxies is centered on the direction of large-scale galactic motion. Nearby aggregations of galaxies such as the Virgo cluster and the Hydra-Centaurus supercluster (*above*) share in the motion, which is thought to result from the gravitational pull of a vast, distant concentration of mass—the Great Attractor. The galaxies that densely populate the center of the image may belong to the Attractor. Ofer Lahav of the University of Cambridge made the map by programming a computer to merge three galaxy catalogues; the dark strip at center is the band of sky obscured by the Milky Way.

© 1987 SCIENTIFIC AMERICAN, INC

the great difficulty of measuring distances to other galaxies accurately. A deeper explanation was also possible. That is, the galaxies themselves might have velocities that depart from the Hubble relation. As the fabric of space expands uniformly, carrying the galaxies apart just as raisins in a cake are carried apart by the expanding dough, the galaxies themselves might be in motion: they could have "peculiar" motions with respect to the expanding universe. Such motions would be most evident in the red shifts of nearby galaxies, whose Hubble velocities are relatively low, and because they might reflect the gravitational effects of an uneven distribution of matter, they could have considerable cosmological significance. Nevertheless, the issue was not explored for a number of decades.

Investigators were occupied instead with a different approach to cosmological questions. By studying galaxies at ever greater distances, observers hoped to push their red-shift measurements back in time. The red shift of a distant galaxy, after all, betrays not its present velocity but its velocity millions or billions of years ago, when the light left the galaxy. Comparing the ancient rate of expansion with the rate derived from nearer galaxies, it was reasoned, would show whether the universe will continue to expand, slow to a stop or recollapse.

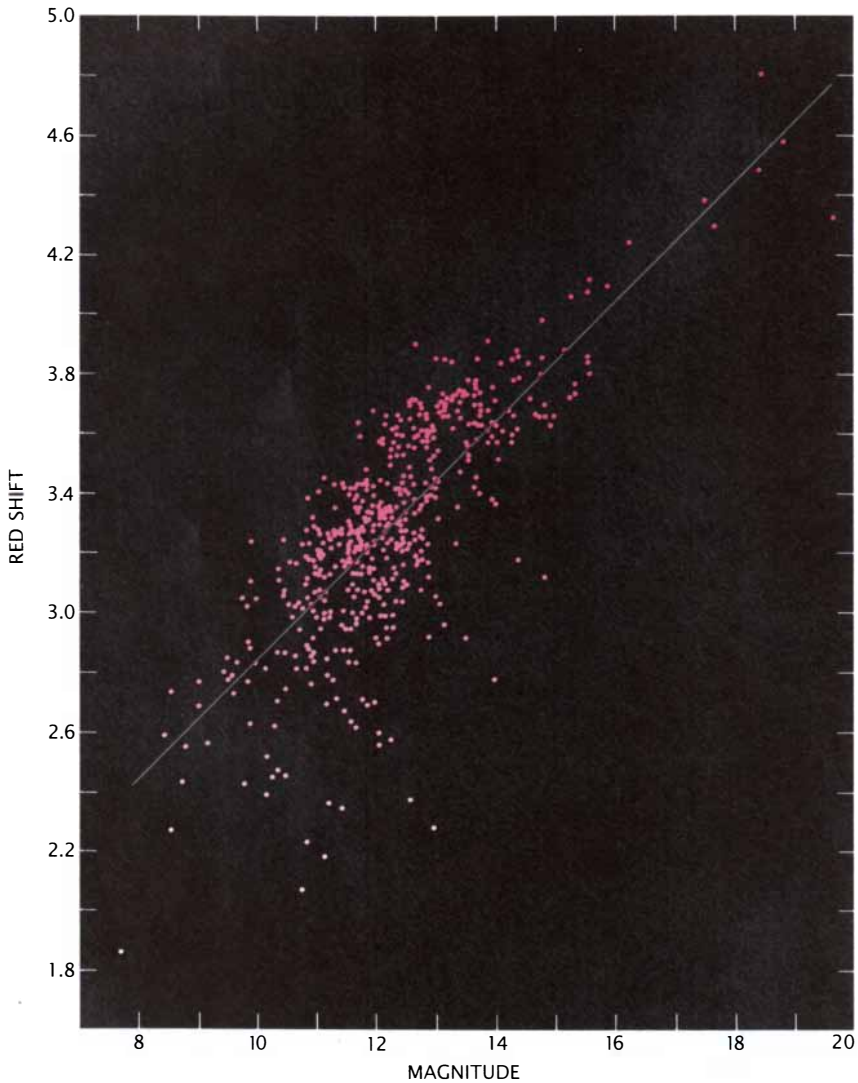
For 40 years investigators worked to wring an answer to this profound cosmological question from distant red-shift measurements, until it became clear that the strategy's fundamental assumption was unwarranted. Red shift was usually measured for the most luminous galaxy in each observed cluster of galaxies; these brightest galaxies were assumed to have the same intrinsic brightness whatever their distance. Under that



© 1987 SCIENTIFIC AMERICAN, INC

assumption one could judge their distance from their apparent brightness. In the 1970's, however, workers realized that such galaxies may change in brightness with time as their populations of stars evolve. At vast removes of distance and time they may not serve as reliable "standard candles."

The frustration of efforts to foresee the evolution of the universe by measuring red shift at distances of billions of light-years led workers to take a new look at the motions of galaxies lying "nearby"—within a few hundred million light-years. Studies of neighboring galaxies, after all, can answer the same cosmological question as measurements of ancient expansion rates. If, for example, nearby astronomical objects are being pulled through space by the gravitational attraction of a concentration of galaxies, their velocities can reveal the attracting mass. By correlating the mass with the observed excess of galaxies, one may be able to determine the amount of mass associated with an average galaxy, and from that mass calculate the density of the local universe. Provided the volume sampled is representative of the universe as a whole, the result can indicate whether the pull of gravity will be sufficient to halt the cosmological expansion.



HUBBLE DIAGRAM indicates that galaxies are receding from the earth at speeds proportional to their distance, testifying to the uniform expansion of space. The vertical axis shows the red shift of sample galaxies: the Doppler shift of lines in their spectrum as a result of motion away from the earth. The galaxies' magnitude is plotted along the horizontal axis. Each unit of increasing magnitude represents a logarithmic decrease in brightness; magnitude gives a rough indication of distance. This Hubble diagram, published by Milton L. Humason, Nicholas U. Mayall and Allan R. Sandage in 1956, shows a feature that puzzled early investigators: low-magnitude (nearby) galaxies have red shifts that depart from the linear relation. It now appears that "peculiar" motions of galaxies through the expanding universe could account for some of the discrepancies.

When the study of peculiar motions began in earnest in the 1970's, it was generally believed that peculiar velocities must be small in relation to the Hubble expansion field. Allan R. Sandage of the Mount Wilson Observatory, for example, argued persuasively that if peculiar motions of a few hundred kilometers per second in random directions are common, many nearby galaxies would show blue shifts: lines in their spectra would be Doppler-shifted toward the blue end of the spectrum, indicating motion in the direction of the earth. Their peculiar velocities in our direction, that is, would more than compensate for the Hubble expansion. The fact that almost all galaxies show red shifts allowed Sandage to argue that the Hubble flow is a very quiet one, from which individual galaxies vary by less than 100 kilometers per second.

Hence a good deal of skepticism greeted Vera C. Rubin and W. Kent Ford, Jr., of the Carnegie Institution of Washington when they reported in 1975 that our own galaxy has a substantial peculiar velocity: about 500 kilometers per second, measured with respect to a distant frame of galaxies. Rubin and Ford arrived at their result by comparing the red shifts of galaxies in a sample distributed around the sky. As long as the reference galaxies lie at about the same distance from the earth, they reasoned, a systematic difference in their red shifts will reflect the motion of our own galaxy. The task was a tricky one, however: if the reference galaxies on one side of the sky were, on the average, farther away than the galaxies on the other side, the difference in red shifts would simply reflect the Hubble flow and not the Milky Way's own motion.

From among the brightest spiral galaxies Rubin and Ford selected a class of galaxies whose intrinsic brightness was thought to be more or less constant. By choosing galaxies of the same apparent brightness they hoped to gain reference points at the same average distance of several hundred million light-years in every direction. Rather than accept the startling conclusion Rubin and Ford drew from the observed differences in red shift, however, many workers preferred to believe they had failed in their attempt to construct an equi-

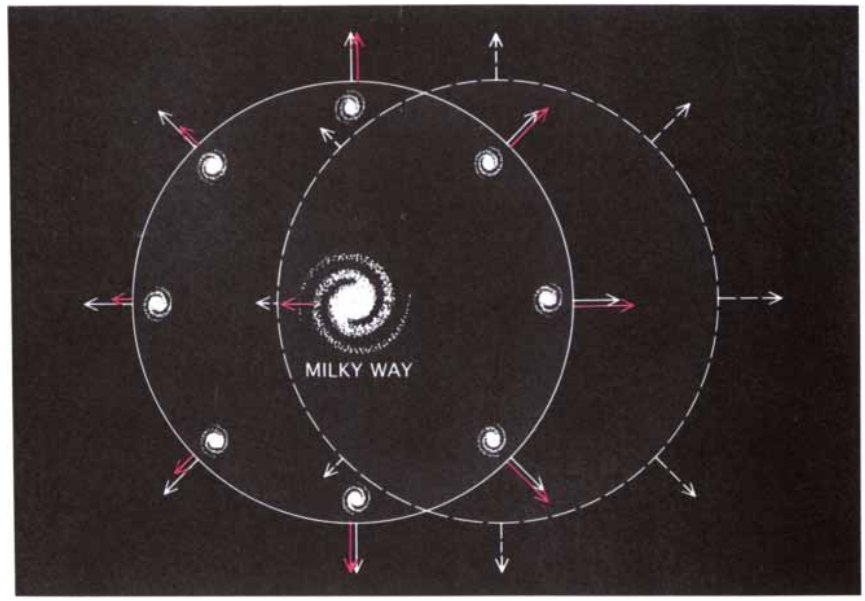
© 1987 SCIENTIFIC AMERICAN, INC

distant reference frame, perhaps because galaxies are distributed unevenly within the range of distance defined by the brightness criterion.

Clearly an acceptable frame of reference is necessary to find the peculiar motion of our own galaxy. In 1977 investigators found that the cosmic background radiation—the faint afterglow of the big bang—can provide the frame. The background radiation was released some 10 million years after the creation, when the plasma that filled the early universe cooled to form hydrogen atoms and the universe became transparent to radiation. Originally given off in the visible and infrared wavelengths, by gas at temperatures of thousands of degrees, the radiation was drastically red-shifted by the rapid expansion of the early universe, so that it reaches the earth today at lower energies, in the form of microwaves. In frequency it matches the radiation emitted by an object at a temperature of 2.7 degrees above absolute zero.

A remarkable property of the microwave background is its uniformity: its temperature is virtually constant across the entire sky. In 1977, however, balloon-borne instruments recorded a small but smooth, sinusoidal variation in the temperature across the sky. In one direction the radiation is a few thousandths of a degree "hotter" than the average; in the opposite direction it is cooler by the same amount. A straightforward explanation was proposed: the radiation is blue-shifted in one direction and red-shifted in the other by the peculiar motion of our galaxy.

Specifically, the pattern suggests the solar system is moving at almost 400 kilometers per second with respect to the microwave background. At the same time the solar system is orbiting the center of the Milky Way, which in turn is drifting toward the Andromeda galaxy, its nearest neighbor. After corrections for these motions, the microwave result implies that the Local Group of galaxies (the Milky Way, Andromeda and several neighboring galaxies) is traveling through space at 600 kilometers per second. It is moving in a direction perpendicular to the line of sight between our solar system and the center of the galaxy, opposite to the direction of galactic rotation and about 27 degrees above the galactic plane. The measurement was soon confirmed, and the notion of a quiet Hubble flow was put to rest. For the first time since Einstein's theory of relativity undermined the notion of an absolute frame of reference, it appeared, one had measured an "absolute" motion: the motion of the Local Group with respect to the universe as a whole.



FIRST CLAIM that the Milky Way is moving through space was made in 1975 by Vera C. Rubin and W. Kent Ford, Jr., of the Carnegie Institution of Washington; their observational strategy is shown. They chose a sample of galaxies surrounding the Milky Way at a distance intended to be roughly constant. According to the Hubble diagram, the recession speed of the sample galaxies should then be the same in all directions (*white arrows*). Actually galaxies on one side of the sky seemed to be moving faster than galaxies on the other side (*red arrows*), suggesting that our galaxy is itself in motion. The finding was equivocal: if the sample galaxies lay closer to the Milky Way in one direction (*broken circle*), the observed difference in the recession velocities could be explained entirely by the Hubble relation between a galaxy's red shift and its distance.

The microwave result vindicated Rubin and Ford's contention that our galaxy is in motion, but the particulars of the findings disagreed: the two velocity vectors pointed in nearly opposite directions. Because the evidence of the microwave background was undeniable, the Rubin and Ford report was filed under "Unexplained results," and workers turned to a new question: What could be responsible for the large peculiar motion of the Local Group?

Conceivably the motion could be a relic of explosive processes during the ancient epoch of galaxy formation. But peculiar velocities generated eons ago would eventually have been lost in the overall expansion of the universe. It is more likely that the impulse responsible for our galaxy's absolute motion is an ongoing one: the gravitational attraction of a concentration of mass. Because most of the visible matter in the universe is included in galaxies, a region of space that contains an excess of them is the most reasonable source of the attraction.

How many extra galaxies would be necessary to accelerate the Local Group to a speed of 600 kilometers per second? Newton's laws of gravitation indicate that the amount of mass needed to produce a given acceleration increases as the square of the distance. Many workers, unwilling to contemplate a large-scale variation in the density of the universe, believe the concentration has to be nearby. At a distance of about 30 million light-years (the distance to the nearest cluster of galaxies) the gravitational pull of many hundreds of extra galaxies, acting over the lifetime of the universe, could generate the observed velocity. At 300 million light-years tens of thousands of galaxies would be needed.

Several groups are now searching for this Great Attractor: the overdensity of galaxies that has the Local Group in its sway. It is not enough to construct maps of the sky; they can reveal clusters of galaxies and superclusters—assemblages of several clusters—but not how much mass they contain or the magnitude of

© 1987 SCIENTIFIC AMERICAN, INC

their gravitational effects. Actually the problem is a tougher one than measuring the peculiar motion of the Local Group alone: to find the Great Attractor one must measure the peculiar motions of many *other* galaxies. Provided the sample is dispersed in a volume of space large enough to include the overdensity itself, the pattern of motions will betray the location of the attracting mass.

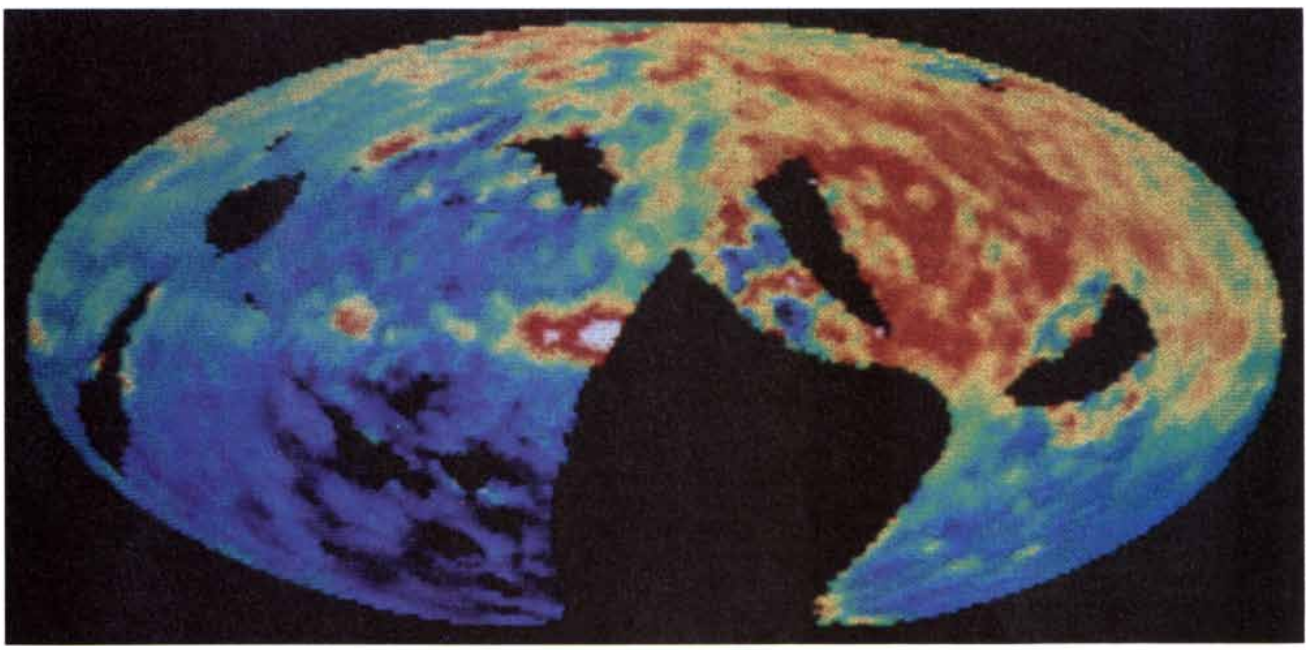
In practice one measures the velocities of galaxies lying at known distances within a large volume of space. After compensating for the Local Group's own peculiar motion of 600 kilometers per second and subtracting the fraction of each galaxy's recession speed due to the Hubble expansion of space, one is left with the galaxy's peculiar velocity—underestimated, in general, because red shift only reveals the component of motion toward or away from the earth. If the peculiar velocities of galaxies within a sphere centered on the Local Group are similar in direction and speed to the Local Group's own motion, it is sensible to conclude that the Great Attractor lies outside the sphere and is pulling on all the galaxies in the region. The task is then to extend the sphere until the outermost galaxies are at rest with respect to the overall Hubble expansion of the universe.

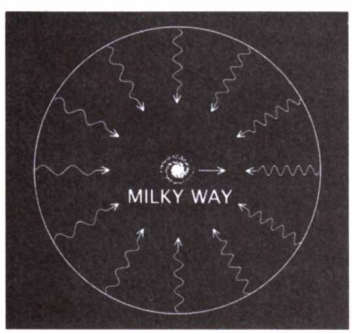
The resulting sphere includes the overdensity responsible for the peculiar motion of the Local Group and other nearby galaxies. In the direction of the overdensity peculiar velocities will increase, as if the galaxies were falling toward the attracting mass. (In fact the distance between the attractor and the "infalling" galaxies may still be increasing because of the Hubble flow; the attractor is only slowing the flow.) Galaxies belonging to the attracting mass itself will be more or less at rest with respect to the microwave background.

Information of two kinds is needed to map the peculiar motions of other galaxies. The first is red shift, which indicates a galaxy's velocity away from the earth. The second is distance, from which one can determine how much of the red shift is accounted for by the expansion of space. Hence the strategy requires a good method for estimating distance, because a galaxy's peculiar velocity may be only a small fraction of its observed velocity.

Instead of taking particular kinds of galaxies as standard candles, assumed to have a constant intrinsic brightness, workers sought a more empirical and testable method for estimating distances. The trick was to find a measurable parameter that bears a fixed relation to the intrinsic brightness of a galaxy and is not affected by its distance. The first such parameter to be applied to the task of measuring peculiar velocities was the rotation speed of spiral galaxies. The correlation between rotation speed and galaxy brightness is known as the Tully-Fisher relation.

The Tully-Fisher relation follows naturally from Newtonian mechanics. The rotation speed of a spiral galaxy is actually a measure of the orbital velocities of stars in the galactic disk. Orbital velocity should depend only on the mass of the galaxy, assuming that all spirals have similar mass densities. If one also assumes





TEMPERATURE OF THE MICROWAVE BACKGROUND was conclusive evidence that our galaxy has an absolute motion. The background is a faint relic of radiation that streamed in all directions through the uniformly hot and dense gas of the early universe. A 360-degree panorama of the microwave sky (*top*), its equator aligned with the plane of the galaxy, shows that this uniform signal is "hotter" by a few thousandths of a degree in one direction (*white*) than in the opposite direction (*dark blue*). The variation is thought to be a Doppler effect resulting from the motion of our galaxy with respect to the universe as a whole (*left*). The map was made by Dale Fixsen, Edward S. Cheng and David T. Wilkinson of Princeton University from data gathered by balloon-borne instruments; the gap at the bottom reflects incomplete data from the Southern Hemisphere.

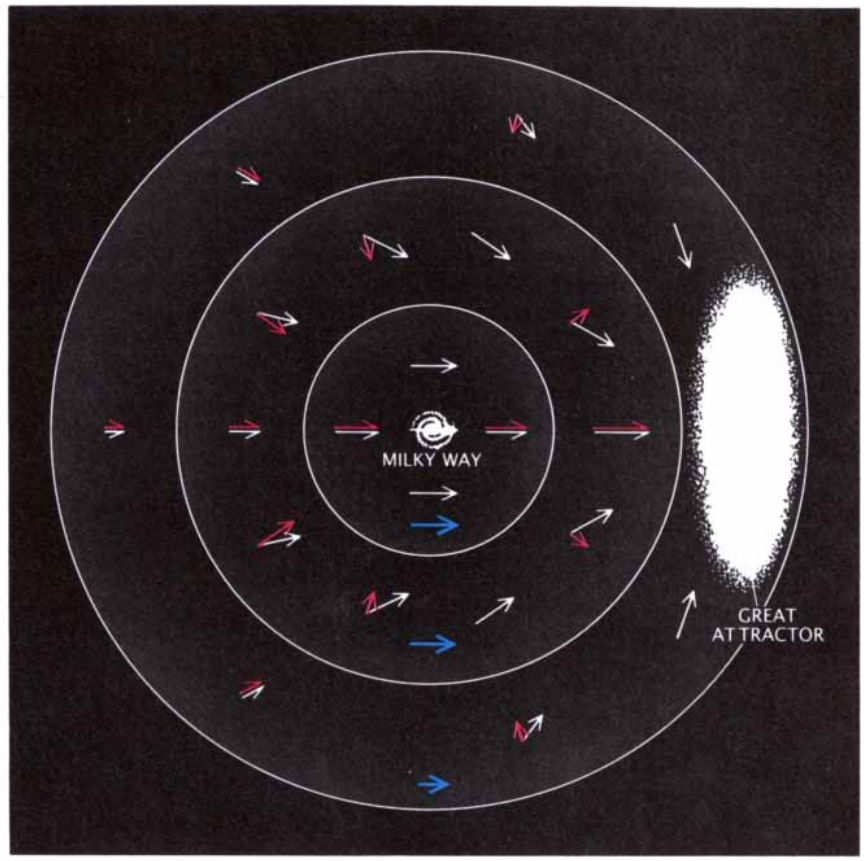
© 1987 SCIENTIFIC AMERICAN, INC

that such galaxies produce the same amount of light for a given amount of mass, then rotation speed can serve as an indication of a galaxy's intrinsic brightness.

The method and its inherent assumptions can be tested by comparing rotation rate with brightness for a large sample of galaxies in a single cluster; because such galaxies lie at about the same distance from our galaxy, their apparent brightness will be proportional to their intrinsic brightness. Rotation rate, like recession speed, can be deduced from a galaxy's spectrum. The difference in the Doppler shift of light or radio waves coming from opposite sides of a rotating galaxy causes a broadening of spectral lines. After a correction is made for the angle between the rotating galaxy and the line of sight, the degree of broadening indicates the rotation speed. Marc Aaronson of the University of Arizona, Jeremy Mould of the California Institute of Technology, John P. Huchra of the Harvard-Smithsonian Center for Astrophysics and their collaborators have found that the rotation speeds and the observed brightnesses of galaxies in the Coma cluster and in several others depart from a strict correlation by no more than 20 percent.

The distance to any other spiral galaxy can be estimated by using cluster samples as benchmarks. After measuring the rotation rate of the galaxy, one refers to the Tully-Fisher relation for, say, the Coma cluster to find the brightness of a reference galaxy that has the same rotation rate. Comparing the brightness of the observed galaxy with that of the reference galaxy gives the distance of the observed galaxy in terms of the cluster. The distance to the Coma cluster itself has not been measured, of course—only its Hubble velocity of 7,200 kilometers per second. It is therefore conventional to express the distance of the observed galaxy in terms of its expected Hubble velocity. Peculiar motion is then simple to compute, by subtracting the galaxy's expected velocity from its actual velocity.

There are several complicating factors. The Tully-Fisher relation suffers from an intrinsic uncertainty of 20 percent for each observed galaxy. Moreover, the reference cluster itself may have a peculiar motion: not all the Coma cluster's velocity of 7,200 kilometers per second may be due to the Hubble flow. To compensate for these sources of error, many galaxies must be studied over a broad area of sky and several different reference clusters must be employed before one can expect statistically significant results.



STRATEGY FOR FINDING THE GREAT ATTRACTOR—the concentration of mass responsible for the peculiar motions of the Milky Way and many other galaxies—is diagrammed. One deduces the absolute motions (*white arrows*) of galaxies in successively larger volumes of space from the component of motion that is measurable directly, based on galactic red shift: peculiar motion toward or away from the earth (*red arrows*). If all the galaxies in a spherical region show similar peculiar velocities, the Great Attractor presumably lies beyond the observed galaxies and is pulling on all of them. By extending the observations until the average peculiar velocity (*blue arrows*) of the most distant galaxies approaches zero one gets a sphere that includes the Great Attractor.

In 1982 Aaronson, Huchra, Mould, Paul L. Schechter of the Mount Wilson and Las Campanas Observatories and R. Brent Tully of the University of Hawaii at Honolulu applied the Tully-Fisher relation in the first effort to identify a cause for the Local Group's peculiar motion. In the local galaxy distribution the most visible feature—and presumably the most massive one—is the Virgo cluster and its entourage of galaxies, named the Local Supercluster by Gerard de Vaucouleurs of the University of Texas at Austin. The center of Virgo, between 40 and 80 million light-years distant, lies about 50 degrees away from the direction of the microwave dipole, which suggested that not all the Local Group's motion could be attributed to the Local Supercluster. An explicit goal of the Aaronson study was to find out just how much of the motion it is responsible for.

The investigators mapped the peculiar velocities of several hundred spiral galaxies at various distances from the center of the Local Supercluster. Stepping out from the Local Group toward Virgo, the observations showed that velocities in that direction increase steadily. On the far side of Virgo the velocity pattern was reversed. The observations thus showed a clear "deceleration pattern" in the supercluster; overdensity is slowing the Hubble expansion in that region of space.

The study provided a convincing demonstration that regions of overdensity can generate peculiar velocities over a broad area. But what did it say about the origin of the Local

© 1987 SCIENTIFIC AMERICAN, INC

Group's motion with respect to the microwave background? The peculiar velocity of the Local Group implies more than 400 kilometers per second of motion in the direction of the Local Supercluster. The Aaronson group's study, like several later ones, found that the Local Group's motion in the direction of Virgo amounts to only about 250 kilometers per second. Aaronson and his colleagues assumed the discrepancy simply reflected measurement errors and random motion of the Milky Way. It could also indicate, however, that the Local Supercluster itself has a velocity of at least 150 kilometers per second with respect to the microwave background. If a mass as large as an entire supercluster could have a significant peculiar velocity, the universe might be lumpy on scales larger than anyone had suspected—an implication that was not taken very seriously at the time.

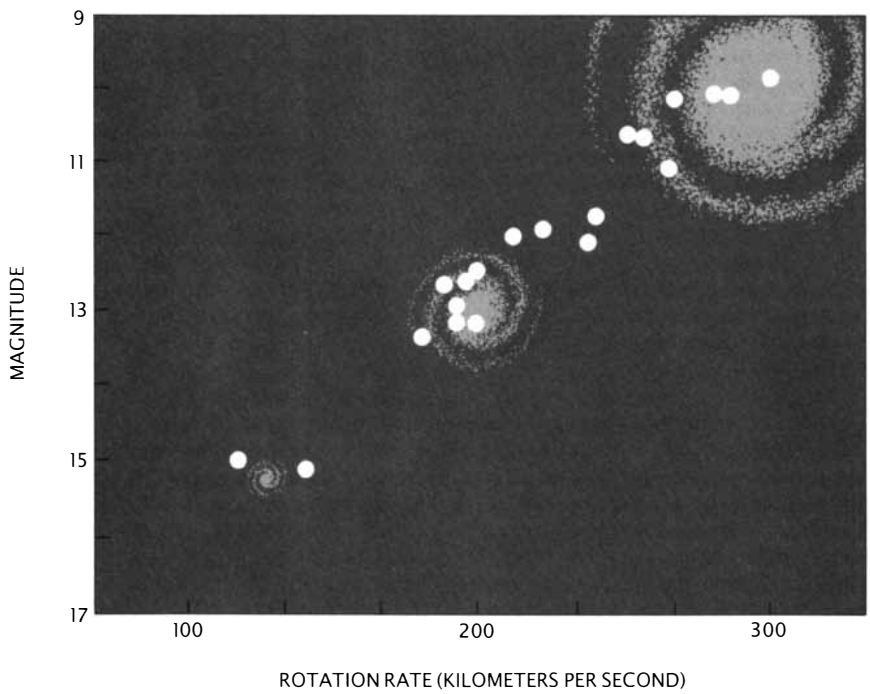
By 1982, then, it was clear that the Local Supercluster does impart a peculiar motion to the Local Group, but it was not at all clear that the "kick" accounts for the Local Group's velocity even in the direction of Virgo. The bulk of the motion, of course, is oriented in quite a different direction and therefore calls for another explanation. Several astronomers noted that if one subtracts a vector representing a motion of 250 kilometers per second in the direction of Virgo from the vector derived from the microwave background, the remaining velocity vector points toward Hydra-Centaurus, the next-closest supercluster. They suggested that this overdensity, which lies about twice as far away as the center of the Local Supercluster, is the major culprit in the pattern of peculiar motions. A disturbing corollary of this attractively simple model is that to account for the observed velocities Hydra-Centaurus must be 10 times as massive as the Local Supercluster—much more massive than its luminosity suggests.

The first test of this explanation was made by David Burstein of Arizona State University, Roger Davies of the National Optical Astronomy Observatories, Sandra M. Faber of the University of California at Santa Cruz, Donald Lynden-Bell of the University of Cambridge, Roberto Terlevich of the Royal Greenwich Observatories, Gary A. Wegner of Dartmouth College and me. The group, dubbed the Seven Samurai by our colleagues, is completing a five-year study of the distances and peculiar motions of about 400 galaxies. Working with telescopes in the U.S., Chile, Australia and South Africa, we have observed galaxies spaced more or less uniformly in every direction, throughout a volume with a diameter of more than 10,000 kilometers per second (several hundred million light-years). That volume encloses the Hydra-Centaurus supercluster, and so the study is capable of showing whether this supercluster is at rest while the Local Group and test galaxies in the direction of Hydra-Centaurus are being drawn toward it.

Our study differs from earlier ones in that it extends to a significantly greater average distance and surveys the sky uniformly. It also differs in that it employs elliptical galaxies, for which we have developed a new technique of estimating distances. Like the Tully-Fisher method for spiral galaxies, the technique is founded on a parameter deduced from the Doppler broadening of spectral lines. Elliptical galaxies do not rotate; instead the broadening results from variations in the line-of-sight velocities of an elliptical galaxy's stars, which follow randomly directed, plunging paths.

Measuring this velocity dispersion is in some ways a simpler task than measuring the rotation rate of a spiral galaxy, since the observer does not have to correct for the angle from which the galaxy is seen or contend with the obscuring dust that fills the arms of a spiral. In observations of elliptical galaxies in clusters we found that the velocity dispersion correlates well with the diameter of the galaxy, defined as the angular diameter of the region having a specific average surface brightness. Like the Tully-Fisher relation, this new relation makes it possible to estimate the distance of a test galaxy with an accuracy of 20 percent; one measures the velocity dispersion and compares the apparent diameter of the galaxy with the diameter of a matching galaxy in a reference cluster.

When we estimated the distances of ellipticals in the region of Hydra-Centaurus and determined how much of their red shifts reflect peculiar motion, we were surprised to find that the sample galaxies in that region of the sky are not at rest. They are themselves moving with respect to the microwave background, per-



TULLY-FISHER RELATION is the relation between luminosity and rotation rate that characterizes spiral galaxies; it serves as a tool for estimating cosmological distances. The relation follows from the fact that both the luminosity of a galaxy and the orbital velocity of the stars in its arms are determined by galaxy mass. The galaxies plotted here belong to the Pisces cluster; the distance to any other spiral galaxy in terms of Pisces can be estimated by comparing the spiral's apparent brightness with that of a Pisces galaxy having the same rotation rate. The graph is from work by Marc Aaronson of the University of Arizona and Jeremy Mould of the California Institute of Technology.

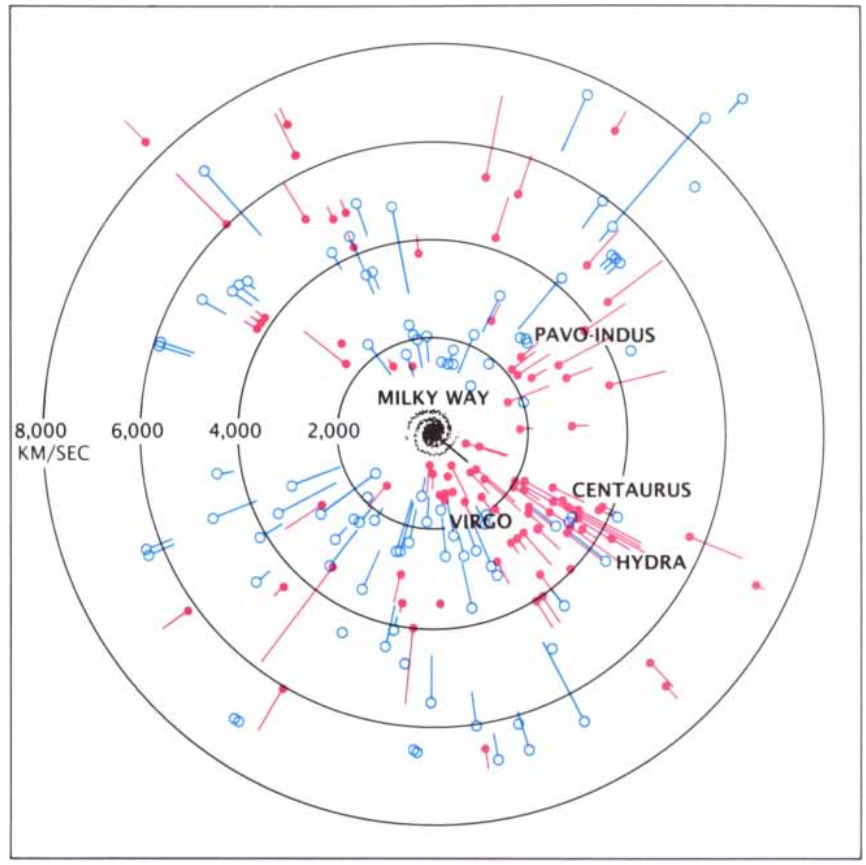
© 1987 SCIENTIFIC AMERICAN, INC

haps even faster than the Local Group and in about the same direction. Instead of being the attractor, it seems that Hydra-Centaurus is itself in the grip of a still more distant and more massive overdensity. Galaxies in the Pavo-Indus Supercluster, which lies at about the same distance as Hydra-Centaurus but about 50 degrees away on the other side of the galactic plane, are also in motion, apparently under the influence of the same distant attractor.

Thus the Local Group's peculiar motion is typical of galaxies over a much larger volume of space than had been supposed. The finding answers Sandage's argument that the paucity of galaxies showing blue shifts is evidence that peculiar motions are rare. Large peculiar motions, on the contrary, are routine, but because they are coherent over huge volumes of space, they do not cause neighboring galaxies to converge on the Milky Way.

Our picture, moreover, is a radical departure from suggestions that the Local Group's motion reflects density fluctuations on a modest scale. Just as the nearest overdensity, the Local Supercluster, turned out not to be the Great Attractor, so it appears that the Hydra-Centaurus supercluster—the next candidate in order of distance—has been found innocent. The Great Attractor seems to be farther away and hence more massive than we had dared to imagine. Indeed, we now have a new candidate. At the limits of our data, at about twice the distance of Hydra-Centaurus but in the same direction, the peculiar velocities of galaxies seem to decline, suggesting that we are approaching the Great Attractor itself.

At that distance the Attractor would have to include several tens of thousands of excess galaxies—about 20 times the excess mass of the Local Supercluster. Efforts are now under way to confirm our result, using the Tully-Fisher relation for spiral galaxies, and also to detect this massive concentration of galaxies directly. Recently I measured the red shifts of about 500 galaxies in the direction of the Great Attractor to get rough indications of their distances. The galaxies are indeed concentrated at about twice the distance of Hydra-Centaurus.



PECULIAR MOTIONS of galaxies out to a distance of several hundred million light-years, corresponding to a recession velocity of more than 8,000 kilometers per second, were charted by the author and his colleagues. The length of the line associated with each galaxy indicates the magnitude of its peculiar motion; red indicates motion away from the earth and blue motion toward the earth. The Milky Way's absolute motion of 600 kilometers per second is also indicated. The data show that galaxies out to great distances, including the Virgo cluster and the Hydra-Centaurus supercluster, are moving in much the same direction as the Milky Way and its neighbors, and about as fast.

Our group and other groups have surveyed the universe out to great distances in search of an "inertial frame" of galaxies: a frame that is at rest with respect to the microwave background. Saying that we have not yet found such a frame is equivalent to saying that our observations do not enclose a representative piece of the universe. We have not yet reached the scale on which the universe becomes homogeneous.

There is some evidence that workers are finally approaching that elusive scale, on which density fluctuations are small and peculiar velocities average to zero. The Aaronson group, employing the Tully-Fisher relation, searched for peculiar motions in spiral galaxies belonging to 10 clusters about twice as distant as the most distant elliptical galaxies in our sample. Within errors, the spiral sample seems to be at rest with respect to the microwave background. Because the galaxy clusters populate only a narrow band in the sky, however, follow-up studies with broader coverage will be needed to confirm the finding.

Wherever the inertial frame turns out to lie, we have already established that even on a scale of hundreds of millions of light-years the universe is made up of overdense regions, as well as regions relatively empty of matter. If our results are verified, they will point to structures many times larger than the largest features that have been mapped directly, such as the cells of the "bubble universe" described in 1986 by Huchra, Valérie de Lapparent and Margaret J. Geller of the Harvard-Smithsonian Center for Astrophysics. Those smaller-scale structures already present a serious challenge to theoretical accounts of events in the early universe, when the density fluctuations would have originated. Our findings amplify the challenge.

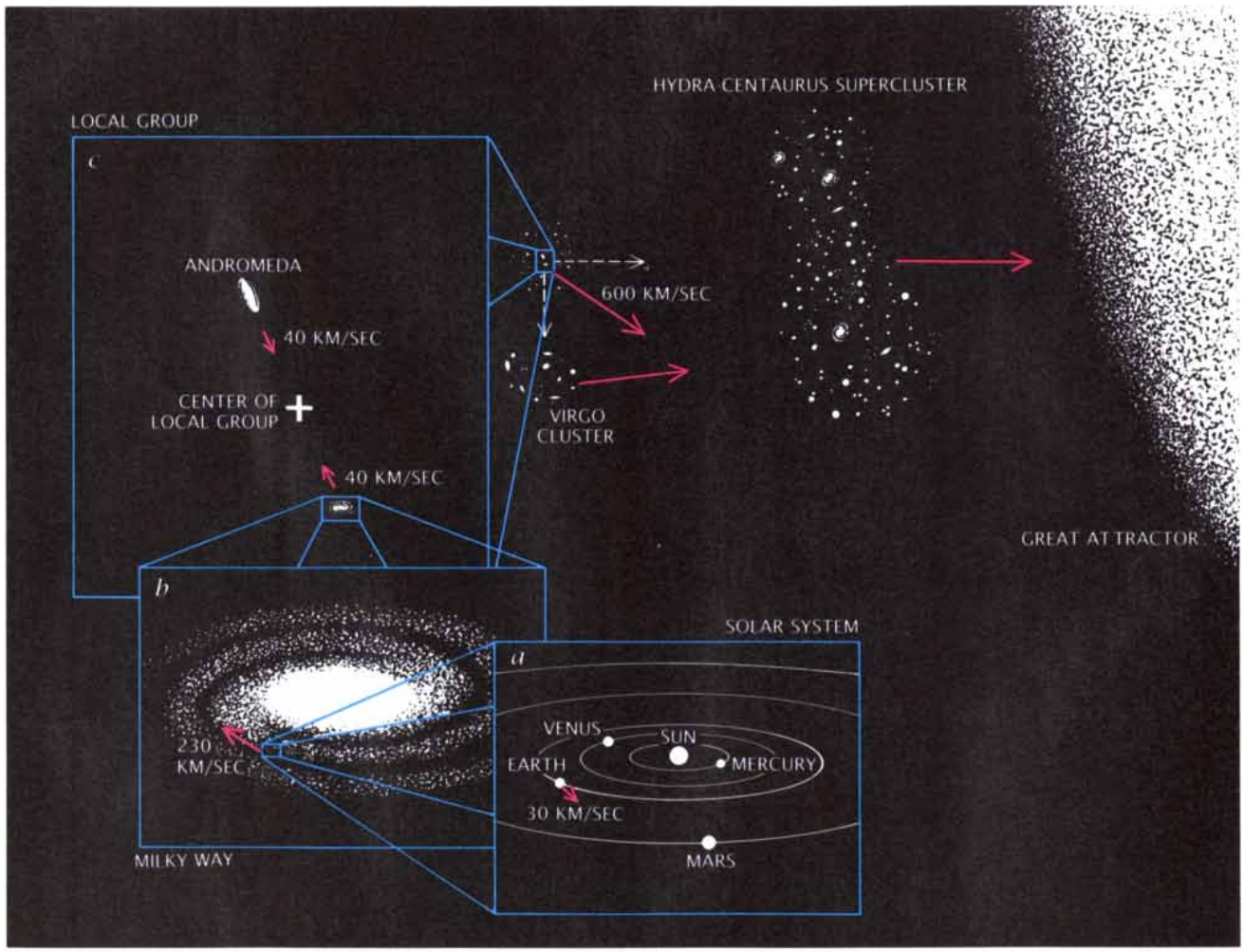
Such theories, after all, must also be able to account for structure on the small scale of galaxies and for the uniformity of the microwave back-

© 1987 SCIENTIFIC AMERICAN, INC

ground, which limits the kinds of inhomogeneities that could have existed in the early universe. Many such accounts assume the existence of some form of dark matter: nonluminous matter that many astrophysicists believe makes up between 90 and 99 percent of the mass of the universe [see "Dark Matter in the Universe," by Lawrence M. Krauss; SCIENTIFIC AMERICAN, December, 1986]. One observed particle, the neutrino, and a number of hypothetical ones have been proposed as the carrier of the missing mass, and models based on various dark-matter candidates predict various spectra of density fluctuations in a big-bang universe as it expanded and cooled. Even the most successful of these theories, however, have not been able to satisfy all the observational constraints simultaneously.

Further observations of peculiar motions will lay the groundwork for a sounder theoretical understanding of the formation and early evolution of the universe. It is even possible that they will extend our theoretical grasp of events at the other end of time. Many theorists now propose that the mass density of the universe is exactly what is needed to cause the cosmological expansion to coast to a stop in the distant future (a proposal that also has powerful implications for processes during the big bang). Estimates of the mass density based on observations of a single galaxy or a group of galaxies consistently give a value too small to slow the expansion effectively. Our study and other attempts to measure the density of matter on the largest scales, however, point to an overall density close to the critical value, and perhaps to a universe in perfect balance.

Exploring the universe is a human enterprise. After this article was written Marc Aaronson, whose work figures prominently here, died in an accident at the four-meter telescope at the Kitt Peak National Observatory. He was a friend and respected colleague, and I shall miss him.



ABSOLUTE MOTION OF THE LOCAL GROUP (the Milky Way and neighboring galaxies) is superposed on smaller-scale relative motions. An earthbound observer must take into account the 30 kilometers per second at which the earth orbits the sun (*a*), the 230 kilometers per second of the solar system's progress around the center of the Milky Way (*b*) and the 40 kilometers per second at which the Milky Way is drawn toward the Andromeda galaxy, its nearest neighbor (*c*). Meanwhile the Local Group is moving at 600 kilometers per second with respect to the fabric of space. The gravitational attraction of the Virgo cluster may account for one component of the motion, but a much larger component is directed toward the Hydra-Centaurus supercluster, which is itself moving. Recent results suggest the Local Group, Virgo and Hydra-Centaurus (and also many other galaxies) are in the gravitational grip of a vast concentration of galaxies at about twice the distance of Hydra-Centaurus: the Great Attractor.

© 1987 SCIENTIFIC AMERICAN, INC

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

The book publishing arm of *Scientific American* for more than 20 years

No Way The Nature of the Impossible

Edited by Philip J. Davis, BROWN UNIVERSITY; and David Park, WILLIAMS COLLEGE

What do we mean by "impossible?" In this unique, mind-opening book, twenty writers—biologists and chemists, poets and mountain climbers, musicians and philosophers—explore the challenges, goals, limits, and illusions of the frontiers of human endeavor. They carry us into original, thought-provoking territory, where we discover that the notion of impossibility is unexpectedly important to our achievements, our visions, our very sense of self.

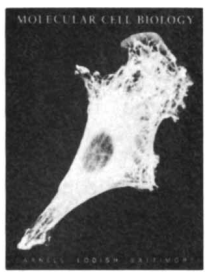
1987, 352 pages, 14 illustrations
Regular Price $17.95 **YOUR PRICE $14.59**

Tilings and Patterns

Branko Grünbaum,
UNIVERSITY OF WASHINGTON; and
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*

1987, 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.; and
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)

1986, 1,187 pages, 1,248 illustrations
Regular Price $42.95 **YOUR PRICE $34.59**

Selling Science
How the Press Covers Science and Technology

Dorothy Nelkin

Whether it's news of AIDS or Star Wars, science and technology make headlines more and more these days. But what does the public really learn from science reporting? And what don't we learn? In this investigation into how science and technology are reported in the U.S. press, Dorothy Nelkin reveals not only the constraints, biases, and myth-making of journalists—but also the surprising public relations ploys of scientists, universities, corporations, and the government. "A healthy dose of reality for everyone concerned about the public's level of information—and misinformation—on today's critical issues...."
—David Perlman, Science Editor
San Francisco Chronicle

1986, 225 pages
Regular Price $16.95 **YOUR PRICE $13.59**

Please send me:

Quan.	ISBN	Title	Special Low Price	Amount
	1813-8	No Way	$14.59	
	1193-1	Tilings and Patterns	$47.99	
	1448-5	Molecular Cell Biology	$34.59	
	1826-X	Selling Science	$13.59	
			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_____

0303

Mail to: **W. H. Freeman and Company**
The Book Publishing Arm of *Scientific American* for More Than 20 Years
4419 West 1980 South, Salt Lake City, Utah 84104

© 1987 SCIENTIFIC AMERICAN, INC

Reverse Transcription

Years ago a group of viruses was found to convert RNA into DNA, reversing the usual sequence of molecular events. Now it appears this process is native to other viruses and many higher organisms

by Harold Varmus

The hereditary material of virtually all living things consists of two strands of DNA twisted into a double helix. The genes encoded in the double helix control cellular activity through two processes: transcription, in which DNA molecules give rise to RNA molecules, followed by translation, in which RNA directs the synthesis of proteins. This DNA-RNA-protein pathway is almost universal.

There are, however, a few exceptions. Some groups of viruses maintain their genetic heritage in RNA rather than in DNA. The genomes of these viruses, which occur in plants, animals and bacteria, can be either single- or double-strand. RNA stores and transmits information in much the same format as DNA does; hence the life cycle of RNA viruses is not too different from that of DNA viruses. Most merely bypass the step in which RNA is transcribed from DNA.

One subclass of RNA viruses, the so-called retroviruses, strays still farther from the conventional path. Retroviruses are capable of making DNA from RNA; in fact, a retrovirus's first maneuver on infecting a cell is to generate a DNA replica of its RNA genome. This viral DNA joins the host cell's DNA and proceeds to direct the production of new viruses.

The retrovirus family is widespread in nature and diverse in pathological potential. Its members include agents that cause leukemia in mice and chickens as well as the virus that causes AIDS in human beings. The distinguishing feature of these viruses, the ability to convert RNA into DNA, is known as reverse transcription.

Reverse transcription was first recognized 17 years ago. For about a decade investigators assumed that the phenomenon was unique to retroviruses, and studies of the process were therefore confined to those viruses and the animal cells they infect. More recently, however, evidence from many quarters has made it clear that reverse transcription takes place under the direction of other viruses and even in the uninfected cells of yeast, insects and mammals. The occurrence of reverse transcription outside the life cycle of a virus may help to explain many diverse and perplexing observations, but it may also raise just as many questions as it answers.

Like all viruses, retroviruses enter a cell and induce it to copy and express their genes, assembling new virus particles with protein coats and RNA cores. Most RNA viruses reproduce only briefly in an infected cell before it dies; retroviral infection, however, is stabler and less injurious to a cell, so that retroviruses are often produced in great numbers without killing the cell host.

In 1970 Howard M. Temin of the University of Wisconsin and David Baltimore of the Massachusetts Institute of Technology independently noted an enzyme in certain RNA viruses that could construct double-strand DNA molecules from single-strand RNA templates [see "RNA-directed DNA Synthesis," by Howard M. Temin; SCIENTIFIC AMERICAN, January, 1972]. The enzyme came to be known as reverse transcriptase and the viruses that contained it were called retroviruses.

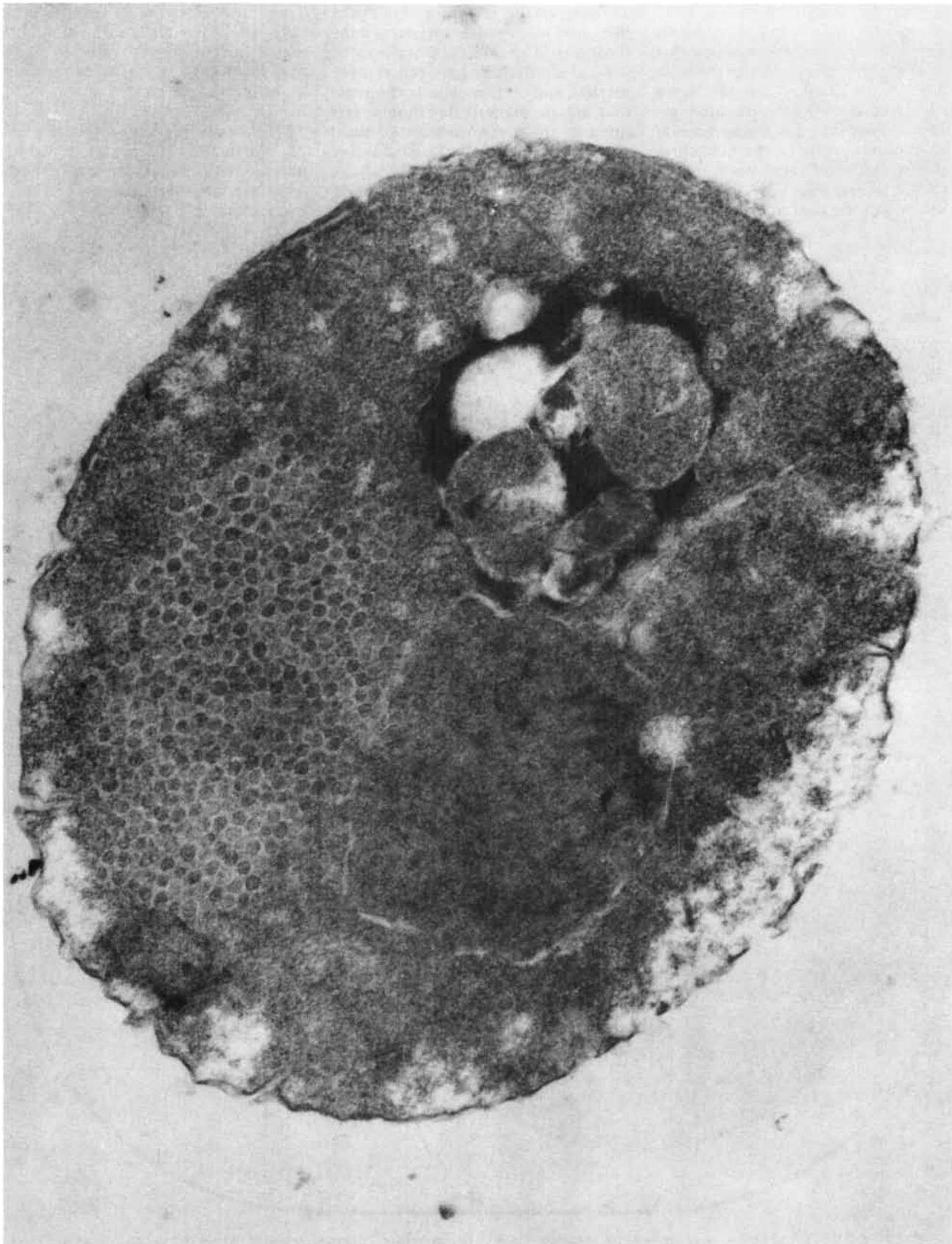
Several years earlier Temin had argued that the stable form of retroviral genes in an infected cell is not RNA but DNA. At the time, however, no enzyme was known to copy RNA into DNA. Temin's theory was vindicated with the discovery of reverse transcriptase. Subsequent studies demonstrated that the DNA form of retroviral genes is inserted into the host cell's chromosomes. The integrated DNA, called the provirus, issues instructions for virus production from its seat in the host genome.

These initial reports stirred great interest in reverse transcriptase. Today its biochemical behavior, its requirements for activity and its response to many drugs have been characterized. It is known, for example, that the curious retroviral enzyme depends on short pieces of RNA as "primers" to help it initiate DNA synthesis and that it cuts up longer pieces of viral RNA to supply some of the primers.

Still, the early studies did not hint at a role for reverse transcription outside the realm of retroviruses. The first clues to a wider scope came in the late 1970's from biochemical studies of the retroviral DNA synthesized by reverse transcriptase. In the laboratory I shared with J. Michael Bishop at the University of California at San Francisco School of Medicine and in John M. Taylor's laboratory at the Institute for Cancer Research in Philadelphia the structure of this DNA was gradually elucidated, with unexpected results. Our experiments revealed that the linear, double-strand forms of retroviral DNA have the same set of sequences at each end; that is, the nucleotides that are the links in DNA chains appeared to be identical over two regions several hundred nucleotides long. These regions are called long terminal repeats, or LTR's.

Initial descriptions of long terminal repeats were based on relatively crude mapping of the few copies of viral DNA found in infected cells. Later several laboratories, including my own, carried out exact determinations confirming that long terminal repeats on each end of a DNA strip were in fact identical nucleotide for nucleotide. It also became apparent

© 1987 SCIENTIFIC AMERICAN, INC



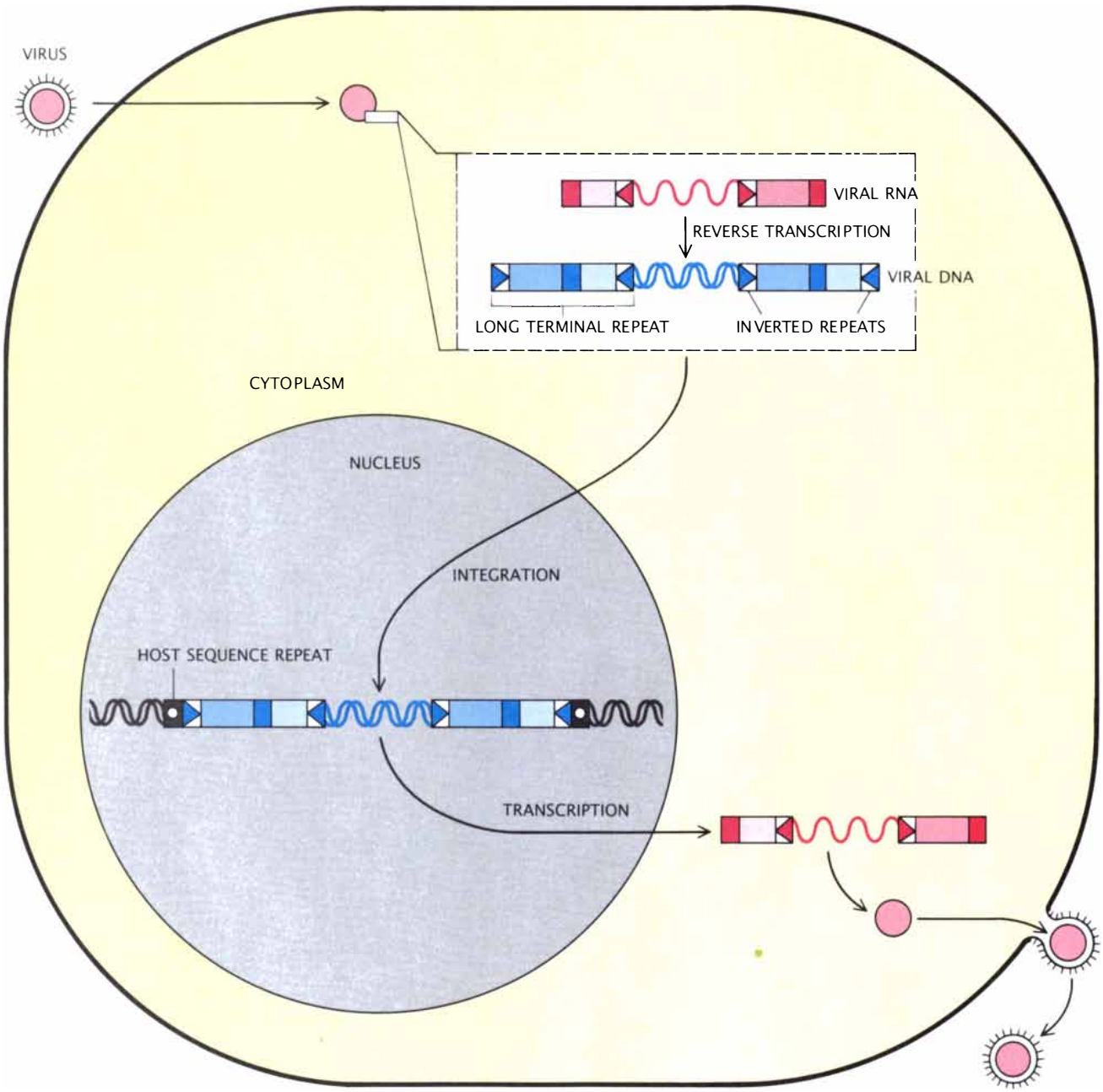
PARTICLES IN YEAST CELL are workstations for reverse transcription. The particles (*left*) resemble those of a retrovirus but are generated by genetic elements in the cell's own chromosomes. The DNA they make is inserted into the chromosomes.

© 1987 SCIENTIFIC AMERICAN, INC

that some sequences near the edge of each LTR are identical if they are read in opposite directions; these were termed inverted repeats. Further sequence determinations revealed that the long terminal repeats of integrated DNA are flanked by the same short sequence from the host genome. Apparently the nucleotides at the integration site on the host chromosome are duplicated when the viral DNA is inserted.

Although the descriptions of LTR's, inverted repeats and host duplications were important to an understanding of the retrovirus life cycle, they had even more intriguing ramifications. The analyses highlighted several similarities between retroviral DNA and certain genetic elements that are permanent but mobile features of the chromosomes of most or all organisms. Called transposable elements, or transposons, they have been described in bacteria, yeast, plants, insects and vertebrates; they are notable for their tendency to switch position in a cell's genome. Transposable elements usually possess "special" ends of some kind and many have LTR's, inverted repeats and short duplications of sequences at the integration site.

The newly perceived relation between proviruses and transposable elements suggested that reverse transcription might be part of the mechanism by which some of the



LIFE CYCLE OF RETROVIRUS is characterized by reverse transcription from RNA (*red*) to DNA (*blue*). The retroviral DNA, which includes genes for several proteins, is flanked by identical sequences known as long terminal repeats; within these are the so-called inverted repeats. When retroviral genes are integrated into the host DNA (*gray*), the host sequence at the integration site also gets duplicated. Only some repeat sequences appear in the transcribed RNA, which is packaged in viral proteins.

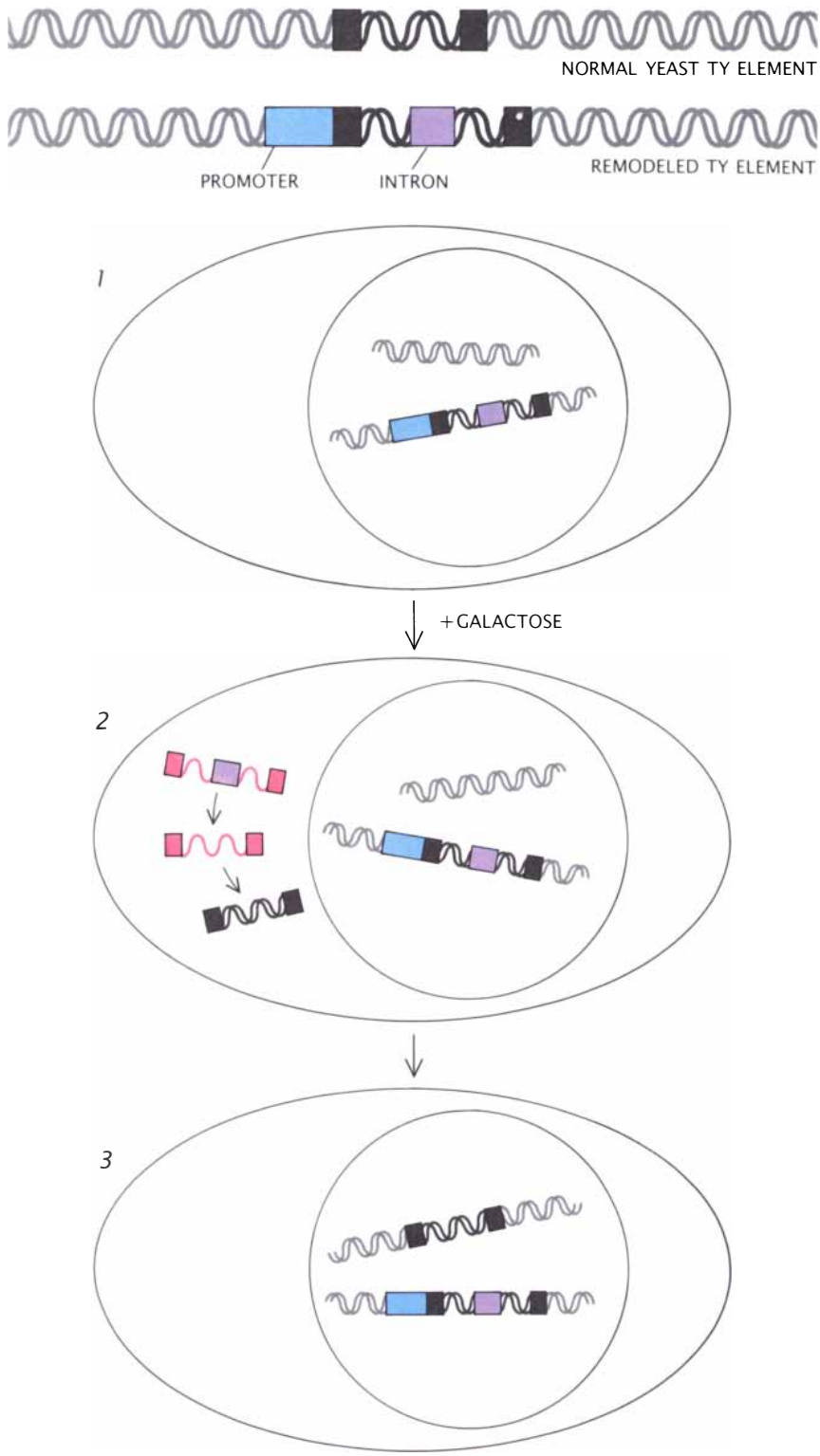
© 1987 SCIENTIFIC AMERICAN, INC

elements move. Other circumstantial evidence implicated reverse transcription. In the transposons of yeast and flies, the nucleotide sequences in the inverted repeats and adjacent to the inner boundaries of the LTR's were very similar to those at analogous positions in proviruses. Sequences at these boundaries are important because in retroviral DNA they are the sites where reverse transcription begins and proviral integration takes place. In addition, computer analysis of the proteins encoded by such transposable elements has revealed remarkable similarities between these gene products and the retroviral proteins responsible for reverse transcription, integration and protein processing [*see illustrations on page 63*].

Although they were provocative, the observations called for more persuasive evidence. Some investigators concentrated on a transposable element known as *copia* in the cultured cells of fruit flies. In 1981 Andrew Flavell and David Ish-Horowicz of the Imperial Cancer Research Fund in London devised a way to detect unintegrated forms of *copia* DNA; they then discovered that the unintegrated *copia* DNA looks much like the unintegrated retrovirus DNA in newly infected cells. Of course, mechanisms other than reverse transcription could have generated the *copia* DNA. A more powerful argument for the involvement of reverse transcription came from functional studies of the yeast transposable element Ty.

The experiments, done by Jef D. Boeke, David J. Garfinkel and Cora A. Styles in Gerald R. Fink's laboratory at M.I.T., exploited the properties of genetic entities known as promoters and introns. Promoters are DNA sequences that regulate the expression of adjacent coding regions. Boeke, Garfinkel and Styles attached to the Ty transposable element a promoter that enhanced the transcription of Ty when yeast cells were exposed to a sugar called galactose. In this way they could indirectly manipulate the transcription of Ty by controlling the amount of galactose in the growth medium: more galactose resulted in more Ty RNA.

Boeke and his colleagues varied the galactose content in the yeast's medium and then examined the modified cells for signs of transposition events. They showed that the appearance of a Ty element at a new position is dependent on the amount of galactose that is present and therefore on the amount of Ty RNA that is



TY EXPERIMENT implicated reverse transcription in the transposition, or movement, of a yeast genetic element called Ty (*dark gray*) from one chromosome to another. Ty was "remodeled" with a galactose promoter (*blue*) that regulates transcription according to the amount of sugar in the yeast's environment. A noncoding DNA sequence called an intron (*purple*), which is excised from transcribed RNA, was also added. Then the yeast carrying the remodeled Ty element was exposed to galactose (*1*). When investigators looked for newly generated copies of the Ty element, they found that the number of transposition events varies with the amount of galactose and that transposed Ty elements have no introns. Hence the Ty element must be transcribed into RNA, processed and reverse-transcribed (*2*) before it is inserted into another chromosome (*3*).

© 1987 SCIENTIFIC AMERICAN, INC

Introducing 4

Cars have come a long way on two wheels. They handle better today than ever before. Their suspension systems are proven on the racecourses of the world. Tire design is its own science.

So automotive engineers started to look behind them. And saw something.

Four-wheel steering would change all this. There would be no delay when you make a turn because all the wheels would turn at the same time. Turning forces would be equal.

Honda engineers thought about this for ten years. They found the answer.

They knew that the rear wheels can play a vital role in the way a car handles. That all four wheels should work together effectively.

It's a reasonable assumption. When you turn your car, the front wheels change direction. They point to where you want to go. The rear wheels remain fixed in a straight line. They have to wait until the car body begins to turn before they change direction.

That means there are two forces acting upon your car in any turn. And always one acts after the other.

That's why your car sways when you change lanes. Why it doesn't turn as sharp when parking. The back end is trying to catch up with the front.

They called it steer-angle dependent four-wheel steering. Here is how it works:

The rear wheels are linked directly to the front wheels by a steering shaft, gears and rods. There are no computers, wiring or electronic black boxes. The Honda system is mechanical and sure. Now comes the amazing part.

How far you turn the steering wheel determines which way the rear wheels turn. And how much they turn.

Turn the steering wheel a little and the rear wheels turn a little in the same direction as the front wheels. Turn it all the way and the rear wheels turn in the opposite direction. There is

© 1987 American Honda Motor Co., Inc

© 1987 SCIENTIFIC AMERICAN, INC

wheel steering.

a big reason for this amazing turn of events. When the rear wheels point in the same direction as the front wheels, there is no delay in the car's turning motion. The car will turn immediately.

You need that for high-speed maneuvers like lane changes. Then you quickly crab over into the next lane. With less body sway.

But low-speed maneuvers are another thing. A car can't turn corners as sharply if the front and rear wheels point in the same direction. Now they need to turn in opposite directions. For example, the front wheels turn right and the rear wheels turn left. The car will then turn in a tighter circle.

Four-wheel steering lets you drive smoothly through gentle curves. Lane changes are more stable. You aren't constantly correcting the steering in a straight line. The car tracks better. That last part is important when you're driving long distances. You work less and arrive more relaxed.

On slippery surfaces, the car is less likely to fishtail. Parking is a breeze. So is getting through a packed parking lot. U turns become u turns.

This year we introduce our latest technological advancement on our most technologically advanced car. The Prelude Si 4WS. It has been refined and redesigned. The hoodline is lower yet. The suspension includes double wishbone on all four wheels. From the racecourses of the world.

There is a new double overhead cam engine with sixteen valves. And a new available automatic transmission.

Honda technology keeps Honda cars far ahead of the others. Often by years. Four-wheel steering promises the turn of the century.

HONDA

The Prelude Si 4WS

© 1987 SCIENTIFIC AMERICAN, INC

generated. Their results suggested that the transposition of Ty DNA requires an RNA intermediate.

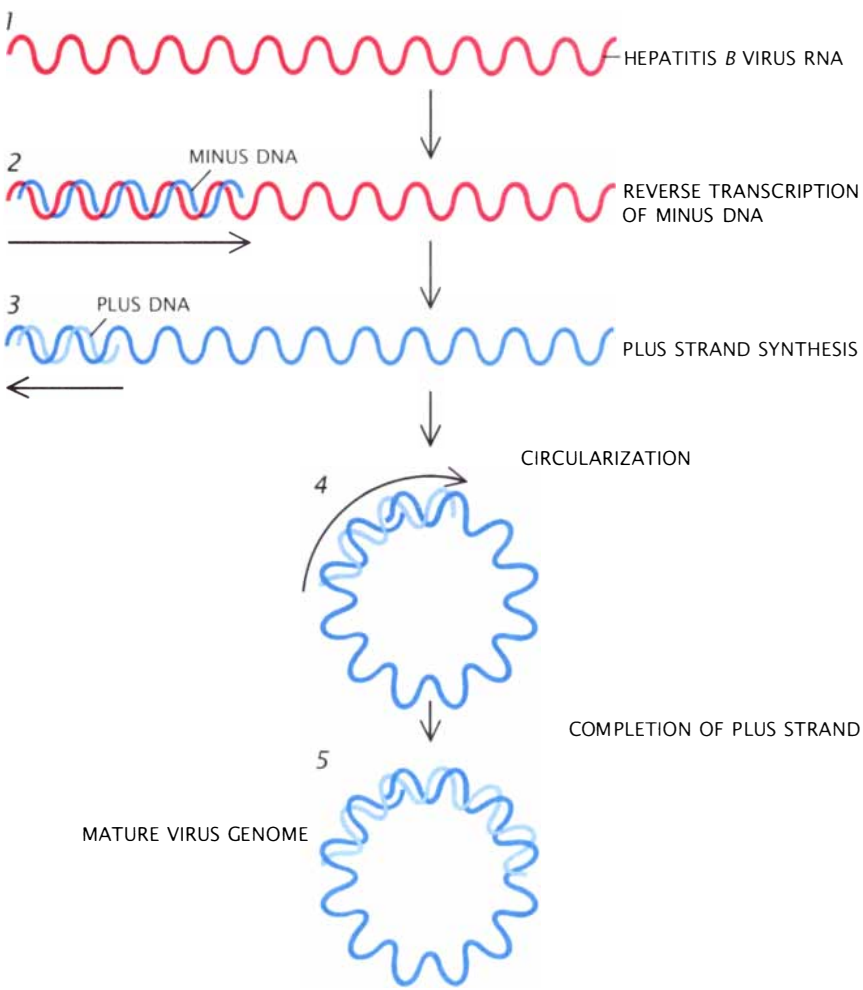
This conclusion was strengthened by experiments in which the M.I.T. investigators inserted an intron into the remodeled Ty element. Introns are stretches of noncoding DNA that are scattered throughout the genes of higher organisms. After a gene has been transcribed into RNA the segments of RNA derived from introns are spliced out of the molecule, leaving a coherent, processed strand of "messenger" RNA. Boeke and his coworkers noted that the intron they introduced into the Ty DNA was absent from new copies of transposed Ty DNA. Somewhere in the course of transposition the intron had been lost. The splicing enzymes that excise introns can cleave only RNA, not DNA; therefore it seems safe to infer that the transposed Ty DNA had been produced from processed RNA rather than from DNA. The link thus established between proviruses and transposable elements such as Ty in yeast suggested a new name for those elements: retrotransposons.

Proviruses have left their mark on other organisms as well. For instance, most normal vertebrate cells harbor retroviral genes; hundreds and even thousands of copies have been found in certain species, including mice and man. The origins of these endogenous proviruses are not known. Most of them cannot produce infectious virus particles and many seem to retain no intact genes at all. A few endogenous proviruses, however, do give rise within cells to particles associated with reverse-transcriptase activity, and these particles, like those produced by yeast and insect retrotransposons, generate new copies of DNA from RNA and insert them indiscriminately into the host chromosomes.

Such promiscuous integration of DNA poses a particular threat to the host cell because mutations often result at the site of integration. The mutation may disrupt a gene's function or change its level of expression and activity. Indeed, many retroviruses induce tumors when their integration prompts an ordinary cellular gene to be expressed at inordinate levels, thereby becoming an "oncogene," or cancer-causing gene.

The persistence in a genome of potentially harmful foreign genetic elements is perplexing. Why are they conserved? Why are they so numerous? Can they confer on the host any selective advantage, or do they survive only by avoiding mutations detrimental to the host? Finally, does the host have any form of protection against the deleterious effects of such mutations?

The preceding questions are intriguing but very difficult to address experimentally. While workers continue to grapple with these puzzles, research done in the past five years has shown that the province of reverse transcription is even larger than had been thought. Several unusual features of DNA viruses and of genes in higher organisms are now being ascribed to the activity of reverse transcription. Although both the viral and the cellular sequences lack some of the cardinal elements of a provirus, the signs of reverse transcription's involvement are quite compelling.



SYNTHESIS OF HEPATITIS *B* VIRUS GENOME demonstrates the role of reverse transcription in the life cycle of a DNA virus. The virus's circular genome consists of two strands of DNA termed plus (*light blue*) and minus (*dark blue*). The longer, minus strand is reverse-transcribed (*1, 2*) from viral RNA (*red*). Synthesis of the plus strand begins near one end of the minus strand (*3*) and continues after circularization (*4*). In the mature genome (*5*) the plus strand is incomplete and the ends of the minus strand overlap.

In 1982 William S. Mason and Jesse Summers of the Institute for Cancer Research proved that hepatitis *B* viruses replicate with the aid of reverse transcription. At first sight any kinship with retroviruses had seemed most unlikely. The hepatitis *B* viruses, the human version of which is an important pathogen, are not RNA viruses: they have a circular, double-strand DNA genome. All cells contain DNA polymerases—enzymes that make DNA from DNA—that viruses can exploit in the course of replication. Why should the hepatitis *B* viruses do anything different?

The first clues came from the structure of the hepatitis *B* genome, whose DNA circle proved to be a peculiar one. Neither of the two DNA strands forms a closed circle, and one strand, the "plus" strand, is shorter than the other, "minus" strand. The plus strand has a fragment of RNA attached to one end whereas the minus

© 1987 SCIENTIFIC AMERICAN, INC

strand has a piece of protein. Mason and his colleagues also noted that minus strands were much more abundant than plus strands in the liver cells the virus infects.

These asymmetries are inconsistent with the prevailing view of DNA replication, which assumes that the two strands of the double helix are duplicated simultaneously and in equal proportion. With reverse transcription, however, there is opportunity for imbalance, because the DNA strands are made sequentially, not simultaneously. The first DNA strand is synthesized from a single-strand RNA template and the second DNA strand is then copied from the first.

Of course, if the DNA genome of a hepatitis *B* virus was regenerated by reverse transcription, it would first have to be transcribed into RNA. Assuming that step occurs, all the pieces fall into place. The structural differences between the two DNA strands could reflect the different templates from which each strand is built; delays in and premature termination of the synthetic process could result in a preponderance of the first strand over the second.

Direct evidence for the involvement of reverse transcription in hepatitis *B* virus replication has not been easy to obtain. Until quite recently the viruses resisted growth in cultured cells, making them difficult to examine with traditional approaches. Mason and Summers' definitive work in 1982 avoided this obstacle by taking virus particles straight from infected tissue. The investigators gathered immature particles of duck hepatitis *B* virus from infected duck livers and incubated the particles with radioactively labeled nucleotides specific to DNA. The labeled nucleotides were incorporated in any DNA synthesized by the particles, so that nascent DNA duplexes and DNA-RNA hybrids could easily be detected by the radioactivity they emitted.

As Mason and Summers had hoped, the isolated virus particles made both plus and minus strands under normal conditions. Then the workers added actinomycin *D,* an antibiotic that stops synthesis from DNA templates. In the presence of actinomycin *D* the copying of plus strands stopped but the copying of minus strands continued, indicating that minus strands are derived from an RNA template rather than a DNA template. This conclusion was confirmed in part by the observation that newly synthesized minus DNA strands are paired with RNA: the density of the hybrid helix is midway between that of pure DNA and that of pure RNA.

Further confirmation of the role of reverse transcription came a year later from another quarter. Takashi Miyata and his colleagues at Kyushu University carried out computer-assisted analyses of hepatitis *B* and retroviral proteins. They uncovered significant similarities between the retroviral enzymes and a protein encoded by a hepatitis *B* virus gene to which no function had been assigned. Most virologists now assume that this protein is a reverse transcriptase; however, it appears to operate somewhat differently from its retroviral counterpart. For instance, it does not equip the DNA with long terminal repeats, which allow retroviral DNA to be integrated into the host's chromosomes.

Soon after that unexpected revolution in thinking about hepatitis *B* viruses, reverse transcription was implicated in more distant territory: a plant virus called cauliflower mosaic virus (CaMV). The evidence supporting its involvement here is less complete but persuasive nevertheless. Like that of the hepatitis *B* viruses, the cauliflower-virus genome is a double-strand DNA circle; sequences near the sites of initiation of CaMV DNA synthesis are similar to those in retroviral genomes to which primers attach. Furthermore, one large cauliflower-virus protein resembles the reverse transcriptases of retroviruses and hepatitis *B* viruses. Roger Hull and Simon N. Covey of the John Innes Institute in England have been able to characterize intermediates in cauliflower-virus replication that are analogous to those seen in retroviral infection; other investigators report reverse-transcriptase activity associated with CaMV particles.

In the examples I have mentioned so far reverse transcription seems to depend on an enzyme provided by an invading virus or the genetic elements affected by it. Now it appears that other less specialized elements of mammalian chromosomes may owe their existence to reverse transcription by enzymes whose origins are uncertain. That surprising discovery emerged tangentially from efforts to catalogue in mouse chromosomes the genes that code for blood proteins called globins.

In vertebrates the globin genes are

MURINE LEUKEMIA VIRUS	TGAAAG . . .	TGGGGG . . .	AGAAAAAGGGGG . . .	CTTTCA
FLY RETROTRANSPOSON	TGTA . . .	TGGCGA . . .	GGAGGGAGA . . .	TACA
YEAST TY ELEMENT	TG . . .	TGGTAG . . .	GGGTGGTA . . .	CA

SEQUENCE SIMILARITIES in the DNA of a leukemia virus, a fly retrotransposon and the Ty element of yeast hint at a common origin. The letters represent the four bases present in DNA: adenine, thymine, guanine and cytosine. Identical bases in the inverted repeats and adjacent regions are in red; color also highlights an area rich in adenine and guanine bases, which are structurally similar. It is difficult to appreciate the inverted repeat (*TG. . .CA*) because only one strand is shown; the sister strand reads *AC. . .GT*.

MURINE LEUKEMIA VIRUS	PQG KNSP . . . R . . . YVDDLLL . . . KK . . . YLG
FLY RETROTRANSPOSON	PFGLKNAP . . . LR . . . YLDDIIV . . . DK . . . FLG
YEAST TY ELEMENT	LYELKCSG . . . VR . . . FVDDMVL . . . EK . . . NLG
HEPATITIS *B* VIRUS	PMG GLSP . . . VR . . . YMDDVVL . . . NK . . . FMG
CAULIFLOWER MOSAIC VIRUS	PFGLKQAP . . . FR . . . YVDDILV . . . KK . . . FLG

PROTEIN SEQUENCES coded for by a retrovirus, by fly and yeast retrotransposons and by two DNA viruses that use reverse transcription also exhibit similarities. Here each letter stands for one of the 20 common amino acids that make up proteins; amino acids that coincide are in red and those that are functionally similar but not identical are in pink. Sequences were derived from a gene encoding the enzyme reverse transcriptase.

© 1987 SCIENTIFIC AMERICAN, INC

usually clustered on two chromosomes, and the coding regions are interrupted by two introns. When they went searching for these genes with molecular probes, Aya and Philip Leder, who were then at the National Institute of Child and Human Development, and Elio F. Vanin, Oliver Smithies and their co-workers at the University of Wisconsin at Madison found an anomalous DNA sequence resembling a globin gene, but with a few important differences. The sequence had no introns and was riddled with mutations that would prevent normal protein synthesis—and it was on a chromosome other than the ones to which the globin genes had been assigned.

One of the models proposed to account for this puzzling phenomenon invoked reverse transcription. Globin RNA that has been processed might be reverse-transcribed and the DNA copy (which would lack introns) might then be inserted at random into the mouse genome. Deprived of the promoter that governs its expression in its normal locale, the new fragment would not be transcribed, and if a gene is not expressed, there is no selective pressure to prevent the accumulation of mutations in the gene sequence.

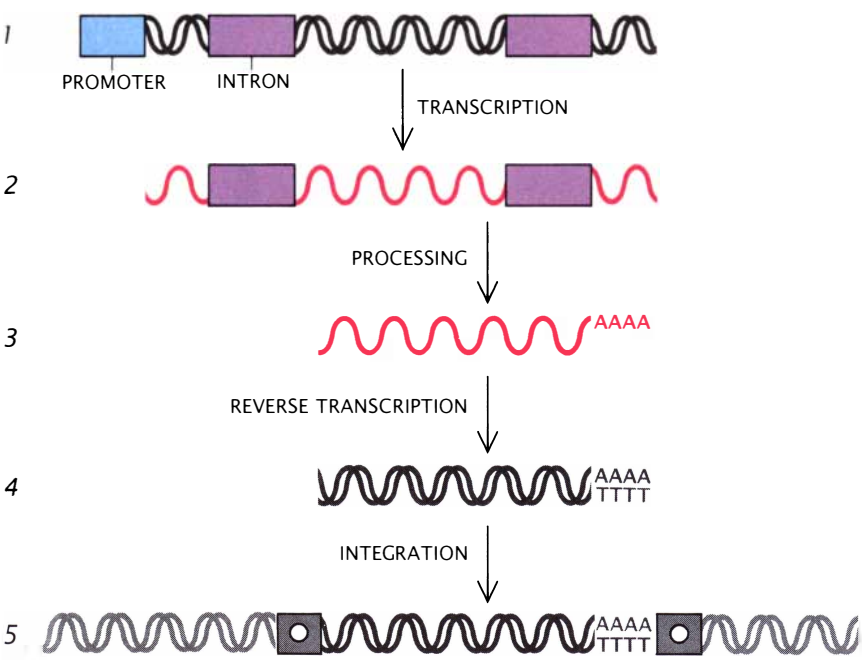
This model took firm hold over the next few years as hundreds of additional examples of "processed" or "retropseudogenes" came to light. The new examples brought more telling evidence for reverse transcription. For instance, some of the globin retropseudogenes have at one end a stretch of DNA in which a particular nucleotide, adenylate monophosphate, is repeated over and over. Such a uniform sequence is not present in globin DNA, but it is represented on globin messenger RNA in the form of a "poly-*A* tail" that is added during processing. Hence the tail in the retropseudogene is a particularly strong indication that it is derived from RNA. Some retropseudogenes also exhibit the host-sequence duplications characteristic of proviruses and retrotransposons.

The recognition of retropseudogenes has established a set of criteria by which other products of reverse transcription may be discovered. Over the past five years several classes of native cellular sequences have been found to display the poly-*A* tail, the duplication of host sequences and the random dispersion among chromosomes that are the hallmarks of RNA-derived DNA. Unlike the globin retropseudogene, most of these new "interspersed elements" seem to have no familiar coding region; a few, however, appear to encode part or all of the reverse-transcriptase enzyme. In a single genome there may be hundreds or thousands of such sequences. Indeed, the aggregated products of reverse transcription—including retrotransposons, endogenous proviruses, retropseudogenes and other interspersed elements—may account for as much as 10 percent of the human genome.

If reverse transcription does in fact figure so largely in the formation of mammalian-cell genes, many questions need to be answered. What is the source of the reverse transcriptase that creates retropseudogenes and interspersed elements? What enzymes select and prepare cellular RNA for reverse transcription, and what enzymes help the new DNA to integrate? How often does reverse transcription occur? Why should the process be part of a cell's functional program at all?

That last question speaks to perhaps the most speculative aspect of research on reverse transcription: the search for the evolutionary origins of genetic information. Some biologists contend that RNA was the first significant molecule of heredity and that DNA is a more recent development [see "RNA," by James E. Darnell, Jr.; SCIENTIFIC AMERICAN, October, 1985]. RNA's versatility supports their contention: it stores information and is replicated in much the same manner as DNA, but unlike DNA it can directly coordinate the synthesis of proteins and can even behave like an enzyme [see "RNA as an Enzyme," by Thomas R. Cech; SCIENTIFIC AMERICAN, November, 1986].

If RNA passed its privileged role to DNA at some point in the history of life, there must have been a primordial mechanism for transferring information from RNA to DNA. Reverse transcription may have been that mechanism. Hence the process that was once considered the esoteric feat of a small band of viruses may now help biologists to retrace the historical path that has produced DNA-dominated forms of life.



MODEL FOR GLOBIN-GENE TRANSPOSITION attempts to account for anomalous versions of the genes giving rise to a blood protein called globin. In mice the "globinlike" genes cannot be expressed because they lack promoters (*blue*); they also do not have introns (*purple*), and they show up on different chromosomes. Ordinarily globin genes (*1*) are transcribed into RNA (*2*), then RNA processing removes introns and adds a "tail" of multiple adenine bases (*3*). The transposition model explains the anomalous globin genes by postulating that the processed-RNA template is reverse-transcribed (*4*) and that the resulting DNA is integrated at random into the mouse genome (*5*). The model is confirmed by the fact that globinlike genes often have an adenine-thymine (*A, T*) tail.

© 1987 SCIENTIFIC AMERICAN, INC

Geneva World Capital of Telecommunications

5th World Telecommunication Exhibition & Forum

20-27 October 1987, Geneva

Where telecommunication leaders, manufacturers, service providers, users, economists and investors from the 162 countries Members of the Union meet in the world's most important and prestigious multifaceted quadrennial communications event covering the fields of national, regional and world communications networks, terrestrial and space communications, public, business, broadcasting and radiocommunication services. Join the decision makers. Come to the 1987 Geneva summit.

Mail to: International Telecommunication Union – ITU – TELECOM Secretariat, Place des Nations - 1211 GENEVA 20 Switzerland

Name/Company/Organization: _____

Address: _____

Please send me information on TELECOM 87 _____ ☐

I wish to participate in the World Telecommunication Forum, please send more detailed information* _____ ☐

Please send me an order form for the Books of Speakers' Papers of Forum 87 _____ ☐

Date: _____ *Signature:* _____

ITU

****FORUM 87*** *19-27 October 1987:* Part I — *Policy Issues of Telecommunications;* Part II — *Technical Symposium;* Part III — *Legal Symposium;* Part IV — *Economic & Financial Issues;* Part V — *Regional Networks.*
Registration Fee for All Parts I-V: 1200 Swiss Francs – before 15 September 1987.

© 1987 SCIENTIFIC AMERICAN, INC

Electrides

In this new class of crystalline materials, trapped electrons take the place of negatively charged atoms. The electrons interact in complex ways, leading to unusual optical and electronic properties

by James L. Dye

One of the simplest and best-known kinds of crystal is the ionic salt, of which a typical example is sodium chloride, or ordinary table salt. The fundamental components of an ionic salt are ions: atoms or molecules that have become electrically charged by gaining or losing one or more electrons. In forming sodium chloride, for example, sodium atoms give up an electron (thereby becoming positively charged) and chlorine atoms gain an electron (thereby becoming negatively charged). The ions are attracted to one another by their opposite charges, and they stack together compactly with crystalline regularity, like tightly packed spheres.

My colleagues and I at Michigan State University have now played a variation on this simple structure, and in doing so we have created a new kind of crystalline matter. In our crystals, which we call electrides, the anions (negative ions) are completely replaced by electrons. The electrons are trapped in cavities within a framework of regularly stacked cations (positive ions). It is easy to trap electrons at defects in ordinary ionic solids, but electrides are the first examples of ionic salts in which all the anionic sites are occupied solely by electrons.

Electrons can in some ways be viewed as the simplest anions possible. Unlike normal anions, however, electrons do not behave as if they were simple charged spheres. In particular, because of their low mass and their quantum-mechanical tendency to interact with one another over great distances, electrons cannot be "pinned down" to any one location as normal anions can. One cannot expect a trapped electron to spend all its time within the cavity; rather, it will wander close to and among the atoms lining the cavity and interact with electrons in nearby cavities, perhaps changing places with them. Electrides therefore have a number of unusual optical, electronic and magnetic properties.

The properties of an electride depend largely on the distance between the cavities that hold trapped electrons and the nature of the channels that connect the cavities. When the trapped electrons are far apart, they do not interact strongly, and so they behave somewhat like a periodic array of isolated negative charges. When they are closer together, they begin to display the properties quantum mechanics predicts for large ensembles of identical particles. When they are still closer, the ensemble properties dominate and the electrons "delocalize": they are no longer tightly bound within individual cavities but instead are more or less free to pass through the spaces within the lattice of positive ions.

Electrides thus provide a fascinating opportunity to study the interactions among electrons in solids. By synthesizing electrides from a variety of materials, one can vary the geometry of the anionic cavities and their relation to the surrounding cations. The resulting properties may make it possible for electrides to become a basis for economically useful new materials and devices.

Related Materials: the Alkalides

Our research on electrides came about as a direct consequence of earlier work we had done to create other novel crystals that we call alkalides. The alkalides have a structure very similar to that of the electrides. They are based on the alkali metals: lithium, sodium, potassium, rubidium, cesium and francium (designated Li, Na, K, Rb, Cs and Fr respectively). Atoms of alkali metals have a strong tendency to give up an electron, forming positive ions. We found that in certain circumstances some alkali atoms can be induced instead to gain an electron, thereby forming a negative ion [see "Anions of the Alkali Metals," by James L. Dye; SCIENTIFIC AMERICAN, July, 1977]. An alkalide, then, is an ionic salt in which both the cations and the anions are derived from alkali metals.

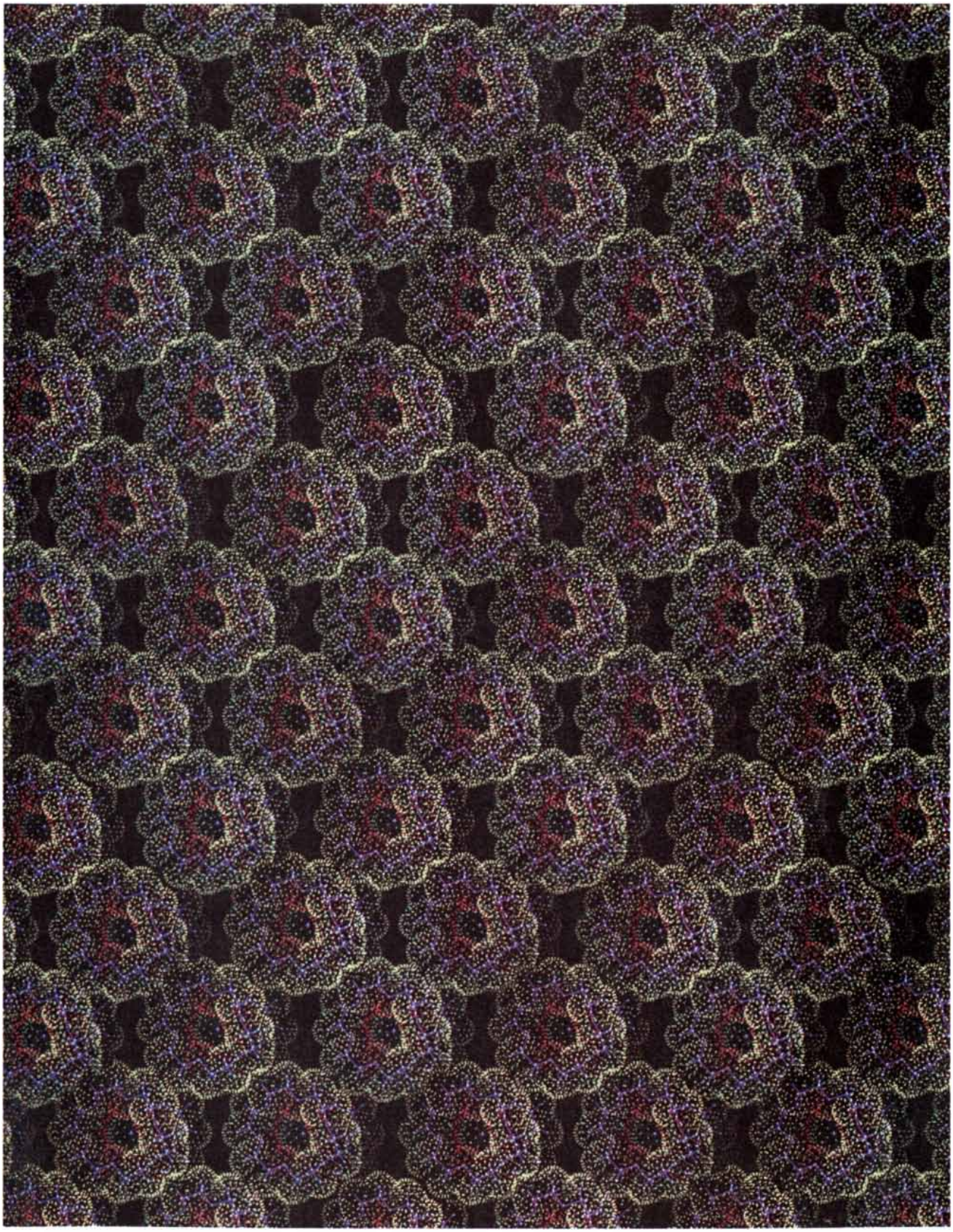
Alkalides require a third component as well. The anions of alkali metals are highly reactive; if alkali anions and alkali cations were simply placed close together, the anions would donate electrons to the cations, thereby neutralizing both kinds of ion and producing a normal alkali metal. To forestall such reactions, we effectively isolated the positively charged ions inside cages: neutral molecules that are not very reactive.

The cages we used were of two types. One type, known as the crown ethers, was developed in 1967 by Charles J. Pedersen of E. I. du Pont de Nemours & Company. The other type, called the cryptands, was developed two years later by Jean-Marie Lehn and his co-workers at the University of Strasbourg.

Crown ethers, as their name suggests, are shaped like crowns, with a ring made up of carbon atoms (each with two hydrogen atoms attached) and crenelations made of oxygen atoms [*see illustration on page 68*]. An alkali cation that is larger than the hole in the center of the crown ether can be caged between two crown ethers: the cation might sit above the center of one crown, surrounded by the crenelations, and the second crown might then fit upside down over the first, like a cap. The two crowns interlock, forming a compact cylinder with the cation at its center and the crown rings at both ends.

Cryptands take the form of a

© 1987 SCIENTIFIC AMERICAN, INC



CRYSTAL STRUCTURE of an electride reveals features that underlie the unusual properties of this new kind of crystalline matter. In this computer image dots define the surfaces of atoms. Each nearly round cluster of atoms represents one side of a "sandwich" of two ringlike structures called crown ethers, which are made up of a carbon backbone (*blue skeleton*) attached to oxygen (*red*) and hydrogen (*yellow*) atoms. In the center of each sandwich is a cesium atom (*green*) that has given up an electron and hence is positively charged. Spaces among the structures are channels connecting cavities in which electrons are trapped.

© 1987 SCIENTIFIC AMERICAN, INC

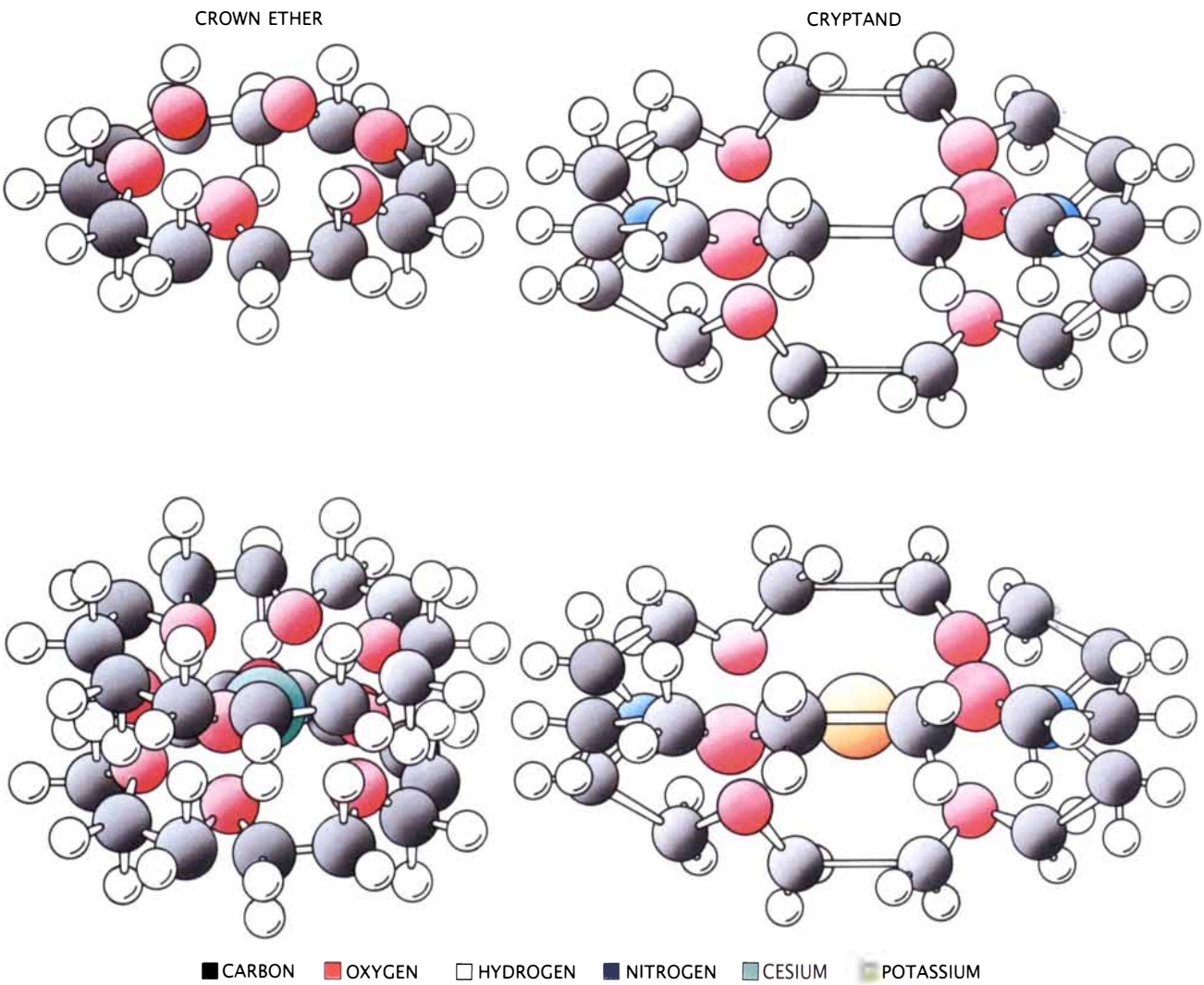
"crypt," or vault. A cryptand molecule consists of two nitrogen atoms connected by three long, curved strands made of carbon, hydrogen and oxygen. The strands act like the bars of a cage: an alkali cation can become trapped within the cage, where it is effectively shielded from any anions in the surrounding medium.

The complexes that are formed when crown ether or cryptand molecules (generally known as complexants) enclose positively charged alkali ions are the structural basis of alkalides (and of electrides). A typical example is the alkalide made of a cesium cation, two crown ethers of the type called 18-crown-6 (because each has a ring structure consisting of 18 atoms, six of which are oxygen atoms) and a sodide anion (a negatively charged atom of sodium). The crown ethers sandwich the cesium ions to form cylinders about 10 angstrom units in diameter and eight angstroms tall (one angstrom is 10^{-8} centimeter, roughly the radius of a hydrogen atom). The cylinders pack very efficiently, filling most of the available space. The cavities left for the anions are shaped rather like an underinflated football measuring about 4.5 by 4.5 by 7.0 angstroms. Each complexed cation is surrounded by eight anions roughly defining the corners of a distorted cube; each anion in turn is surrounded by eight complexed cations.

Synthesizing Electrides

Among the gold-colored crystals in the first batches of crystalline alkalides we synthesized, we occasionally noticed certain dark blue—almost black—solids. They differed from the alkalides not only in color but also in their unusual magnetic properties—properties that made us suspect these solids were a salt that was similar to the alkalides but contained trapped electrons (donated by the metal atoms when they formed cations) in the place of alkali metal anions. In other words, we suspected that we had accidentally synthesized electrides.

This led us to try synthesizing electrides deliberately. Unfortunately it has been extremely difficult to develop controlled techniques for making solids that can be identified unambiguously as electrides. The chief problem is that trapped electrons are chemically "hot" and it is therefore difficult to prevent electrides from



COMPLEXANT MOLECULES serve in electrides to enclose cations (positively charged atoms), preventing them from recombining with electrons. A crown ether (*top left*) is a ring of carbon and hydrogen atoms with crenelations made of oxygen atoms. Two crowns envelop a cation in a cylindrical sandwich (*bottom left*). Cryptands (*top right*) are shaped like a cage with three bars: chains of carbon, oxygen and hydrogen atoms connecting two nitrogen atoms. A cation can fit neatly in the cage (*bottom right*).

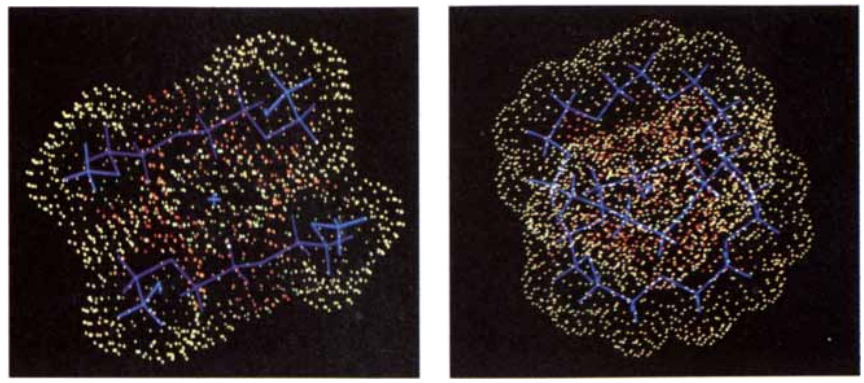
© 1987 SCIENTIFIC AMERICAN, INC

self-destructing: the electrons tend to react with the complexant molecules or with the solvent in which the complexant and the alkali metal are dissolved, decomposing them.

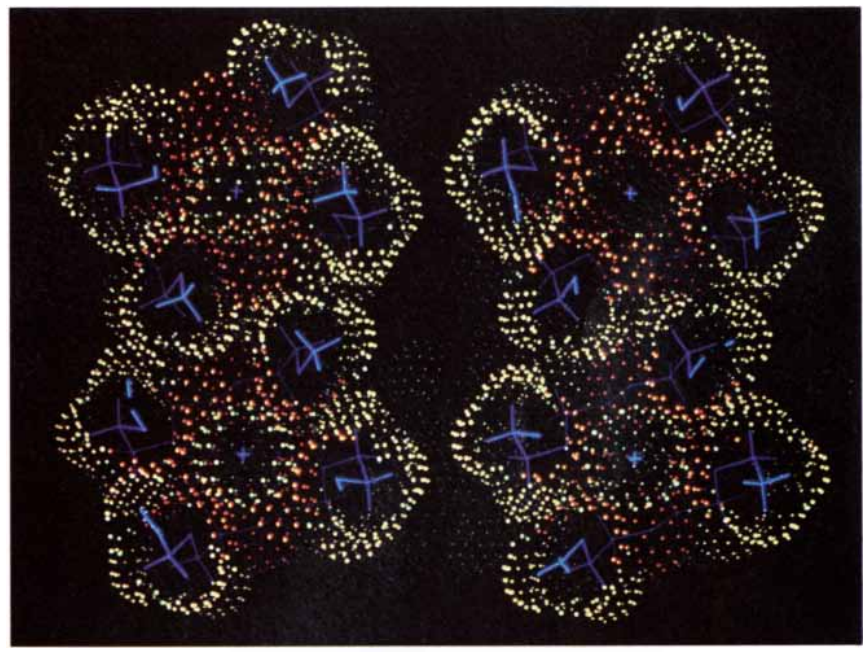
The decomposition reactions are chain reactions (that is, the products of each reaction serve to induce one or more later decomposition reactions), and so they proceed rapidly once they have been initiated. As a result, to synthesize electrides one must be particularly careful in purifying and handling the solvents, complexants and alkali metals. An added complication is that all steps in the synthesis and storage of electrides must be done at very low temperatures because the rate of decomposition increases with increasing temperature.

Even so, the first deliberately produced electrides were prepared relatively simply. When alkali metals are dissolved in certain solvents, they tend to give up an electron into the solution, thereby becoming positively charged [see "The Solvated Electron," by James L. Dye; SCIENTIFIC AMERICAN, February, 1967]. Dissolved complexant molecules spontaneously envelop such dissolved alkali cations. We therefore dissolved appropriate amounts of metal and complexant together and then evaporated the solvent. The result was a fine powder or, sometimes, a thin film of electrides. We were able to learn quite a bit about the nature of electrides by analyzing such products. Nevertheless, the fabrication procedure often yielded products containing various contaminants: other kinds of compounds (such as alkalides), excess metal or complexant, and chemicals produced by the decomposition of the solvent or the complexant.

A second method of fabrication, which yields thin electride films of controllable thickness and composition, is called vapor deposition. In this procedure the complexant and the alkali metal are placed in separate containers under high vacuum. By manipulating the temperatures of the containers, it is possible to evaporate the metal and complexant and deposit them simultaneously onto a quartz surface at adjustable rates. A solid-state reaction between the complexant and the metal produces the electride film. This procedure is useful in studying the optical and electrical properties of thin electride films, but it cannot produce large samples of crystalline electrides because the thickness of the deposited film grows by only a few angstroms per second.



CROWN-ETHER SANDWICH, consisting of two crown ethers enclosing a cesium ion, is depicted from two angles. The view at the left is a cutaway through the center of the crown ethers, showing the placement of the ion in the sandwich. The view at the right shows the shape of the sandwich as a whole, as it would be seen in a three-quarter view.



CLOSELY PACKED ARRAY of crown-ether sandwiches defines a cavity in which an electron can be trapped. Channels connecting this cavity to others are at the top and bottom of the image. The view here is roughly orthogonal to the view shown on page 67.

A third, and perhaps the best, method of preparing electrides is known as direct crystallization. The metal and the complexant molecules are put in separate arms of a special reaction vessel [*see illustration on page 75*]. A solvent is introduced into the arm containing the complexant and the apparatus is tilted so that the dissolved complexant pours over the metal. The metal dissolves as molecules of crown ether or cryptand form complexes with metal cations. Then much of the solvent is evaporated and a weaker cosolvent is added. Next the solution is cooled. If the mixture of solvents has the right composition, crystals form spontaneously in the cooling solution. In some cases we have been able to grow relatively large crystals (about a millimeter on an edge) by this technique. After the solvent has been removed the crystalline samples are poured into glass tubes, which are then heat-sealed and stored at about −80 degrees Celsius until they are needed. During the entire operation the temperature of the samples must never exceed about −20 degrees.

Optical Properties

The electrides produced by the first and third of these methods are

© 1987 SCIENTIFIC AMERICAN, INC

usually fine, blue-black powders or black, shiny crystals. Thin films of electride produced by the first or second method appear blue when a light is shined through them: the films absorb light in the red and infrared regions of the spectrum.

Information on optical absorption by electrides can in fact tell quite a lot about the properties of electrons in these compounds. When a substance absorbs a photon, or quantum of light energy, the energy of that photon (which depends on the photon's wavelength) is transferred to an electron in the substance, and the electron is raised from a state of relatively low energy to a state of higher energy. According to the rules of quantum mechanics, electrons in materials can have only certain discrete amounts of energy. The allowed energy levels vary from substance to substance.

In solids the allowed energy levels are spread into bands; each band consists of a narrow range of allowed energy states. According to the Pauli exclusion principle, only two electrons can occupy the same energy state within a given energy band. For a photon to be absorbed it must therefore have just the right energy to raise an electron from one allowed energy band to another band that contains some unoccupied states or from one state within a band to another state within the same band that is not fully occupied.

The wavelengths of light a substance can absorb therefore indicate the differences in energy between the available electron energy states in that substance. The energy differences in turn indicate how tightly bound the electrons are to atoms in the substance or, in the case of electrides, how strong the forces are that trap the electrons within cavities. The more tightly bound an electron is, the greater the amount of energy is that it must absorb in order to be freed.

On the basis of their optical spectra, electrides can be divided into two rough categories: those in which the electrons acting as anions are localized within the cavities formed among the complexed cations and those in which they are delocalized, or free to move through the crystal (although it may be that even the electrons in delocalized electrides are actually weakly trapped rather than completely free).

The anionic electrons in a delocalized electride seem to be much like the electrons in a normal metal. In a metal the valence electrons (the outermost electrons of each atom) are free to roam throughout the lattice of atoms, with the result that metals tend to be good conductors. These so-called conduction electrons are in an energy band—the conduction band—that is only partly filled with electrons. The conduction electrons can therefore absorb photons in a continuous range of wavelengths, because there is no minimum distance between the available energy states within that band. They can even absorb photons that carry a small amount of energy (that is, photons with long wavelengths). Like metals, delocalized electrides can also absorb photons in a range of wavelengths, including relatively long wavelengths.

Localized electrides—which make up the majority of the electrides synthesized so far—have more in common with semiconductors than they do with metals. Just as electrons in localized electrides are bound within anionic cavities, so the electrons in semiconductors are bound to individual atoms. The electrons can be elevated to the conduction band, a state of greater freedom and higher

THIN FILM OF ELECTRIDE coating the inside of a flask appears blue when light is shined through it; this electride absorbs red and infrared light. The optical absorbance of an electride yields clues to how tightly the trapped electrons are held and how much they interact. The electride in this photograph is made from cesium metal and crown-ether molecules called 18-crown-6; its structure is shown in the computer images on preceding pages. The film was made by Mark Kuchenmeister in the author's laboratory.

© 1987 SCIENTIFIC AMERICAN, INC

energy, when they absorb a photon. The difference between the energy of the bound state and the energy of the so-called excited state sets a minimum to the energy such an electron can absorb. Semiconductors therefore do not absorb low-energy light very strongly, but they do absorb higher-energy light.

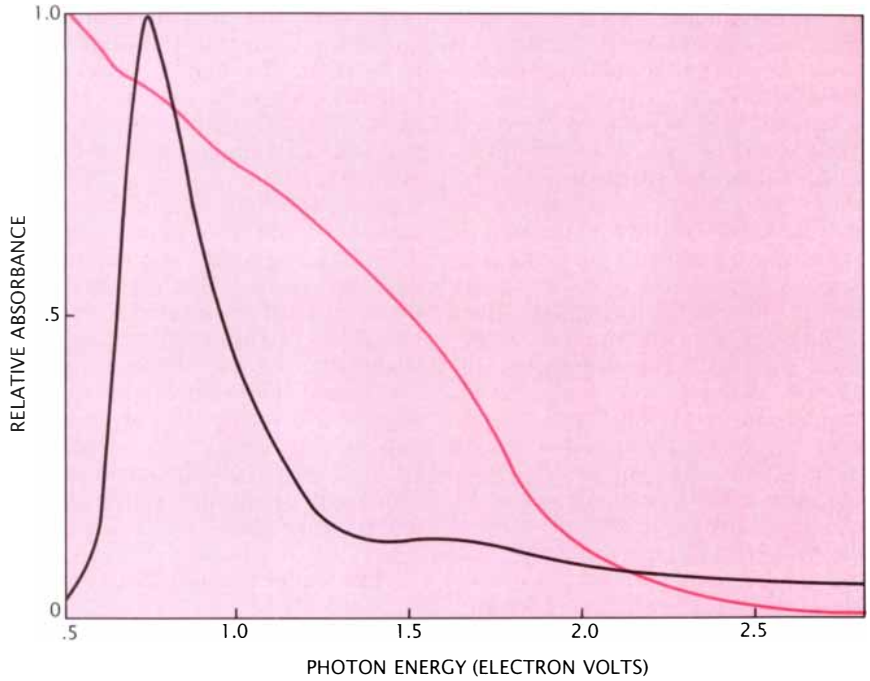
Similarly, localized electrides do not strongly absorb photons having energies below about .5 electron volt (which corresponds to infrared light that has a wavelength of about 2.5 microns), but they do absorb light in shorter wavelengths (that is, higher energies). It is not yet clear whether a photon with an energy of .5 electron volt raises an electron in an electride all the way to the conduction band or only to some weakly bound intermediate state.

Are Some Electrides Metals?

Observations of the optical absorption spectra of electride films thus raised the exciting possibility that electrides of varying composition might display electronic properties spanning the range between the properties of metals and those of semiconductors. Why not settle the question by measuring their electrical conductivity? Unfortunately such a simple proposal faces several inherent complications.

First of all, many nonmetals (including some semiconductors) can conduct electricity quite well. The true test of whether a material is a metal is whether its conductivity rises or falls when the temperature is lowered. Metals become more conductive at lower temperatures. The atomic lattice vibrates less, and the current-carrying electrons therefore experience fewer collisions with lattice atoms. Semiconductors, on the other hand, become less conductive as the temperature is lowered. The reason is that the only electrons free to carry current in semiconductors are those with enough thermal (heat) energy to rise from the bound state to the conduction band. When the temperature is lower, fewer electrons have enough energy to act as conduction electrons. In order to test the electronic properties of an electride, then, it would be necessary to test its conductivity over a range of temperatures.

Another complication arises from the difficulty of growing large electride crystals. The best way to measure the conductivity of a material is to attach four thin wires to a single crystal. A current is passed between two of the probe wires and the voltage difference between the other two wires is measured. It is hard to grow single electride crystals that are large enough for such a test. In addition, electrides' reactivity and sensitivity to heat make the crystals difficult to handle. These problems have so far prevented us from measuring the conductivity of single crystals of electrides.



OPTICAL ABSORBANCE SPECTRA of two electrides manifest differing electronic properties. The electride made from cesium and 18-crown-6 (*black*) tends to absorb photons (quanta of light) that have an energy of about .5 electron volt or higher but not photons of lower energies, indicating that its electrons are strongly localized: each electron is tightly trapped in a cavity, from which it can be released when it absorbs a photon of exactly the right energy. The electride made from lithium and a certain cryptand (*color*) can absorb photons in a broad range of energies, which indicates that its electrons are much more delocalized. That is, they are relatively free to move through the electride crystal, and on absorbing a photon they simply move more energetically.

We have gained useful information, however, by measuring the conductivity of electride powders that have been packed tightly. The resistance caused by boundaries between the minute crystalline grains biases the measurements, but the dependence of a packed powder's conductivity on temperature should not vary too much from that of a single crystal.

To make the measurement we compress a sample of powder between spring-loaded stainless-steel electrodes, which generate about 50 atmospheres of pressure. By varying the temperature of a stream of nitrogen that passes over the sample cell we can gauge how the sample's conductivity varies with temperature.

By means of this technique we have found that the electrides whose optical properties resemble those of semiconductors (those in which the anionic electrons are localized) also show the decrease in conductivity at low temperatures that is characteristic of semiconductors.

An added difficulty arises when we try to measure the conductivity of the electrides whose optical spectra resemble those of metals. The conductivities of packed powders of these compounds are initially much higher than the conductivities of "semiconducting" electrides, but they decrease rapidly with time. A possible explanation is that the stainless-steel plates act as catalysts for the decomposition of the adjacent layer of electrides, creating insulating layers of decomposition products. We have therefore not been able to prove that the "metallic" electrides are indeed metals.

Crystal Structure of an Electride

Why are the anionic electrons in some electrides localized and those in others delocalized? We believe the answer lies partly in the degree to

© 1987 SCIENTIFIC AMERICAN, INC

which the anionic electrons interact with one another, which in turn depends in part on the spatial structure of the crystal.

To test this hypothesis it would make sense to try to learn the physical structure of the various electrides. Unfortunately that is not an easy task either. Once again, part of the problem lies in our not being able to grow large single crystals of electrides. Most of the crystallographic techniques for determining structure, such as X-ray diffraction, require a relatively large single crystal. In addition, X-ray diffraction cannot give any specific information about the positions of trapped electrons; it can only tell the positions of the atoms in a crystal. Nevertheless, from the positions of the alkali atoms and the complexants it should be possible to deduce the shapes, sizes and relative positions of the cavities in which electrons might be trapped.

One electride whose structure has yielded to X-ray crystallography is composed of cesium cations complexed by 18-crown-6 crown ether molecules. This electride's structure is very similar to that of the alkalide made of cesium cations, 18-crown-6 complexants and negatively charged sodium ions, which I have described briefly above. Again a pair of crown ethers encloses each cesium ion to form a cylinder about 10 angstroms in diameter and eight angstroms tall, and the cylinders are packed tightly together. The spaces between the cylinders define the anionic cavities, each of which is surrounded by eight cylinders in a slightly distorted rectangular lattice.

This electride is one of those that acts somewhat like a semiconductor. Its optical properties and conductivity suggest that the anionic electrons are strongly localized, and the physical structure seems to bear the suggestion out. The anionic cavities are rather far apart, and so the trapped electrons probably interact only weakly. Each cavity is connected to six other cavities by channels running between the cylindrical cationic complexes.

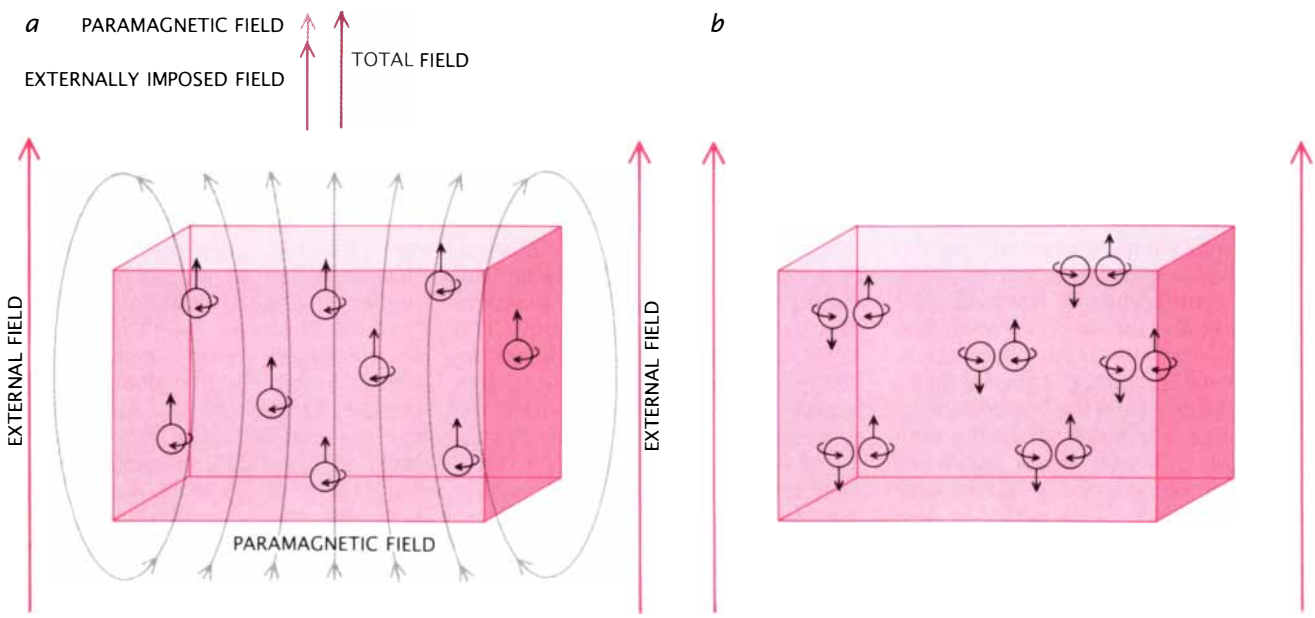
Interactions among Electrons

In the case of other electrides, which have not been analyzed by X-ray crystallography, there are alternative ways to learn about the interactions among anionic electrons. One technique involves measuring how each electride responds to magnetic fields at various temperatures.

The magnetic properties of a solid can be traced largely to the magnetic behavior of individual electrons or groups of electrons within the solid. Every electron generates a small magnetic field like that of a bar magnet. Unlike the field of a bar magnet, the electron's magnetic field follows certain quantum-mechanical rules. For example, when the electron is in an externally created magnetic field, the electron's inherent magnetic field can point only in the same direction as the external field or in the opposite direction. In addition, when two electrons are paired in what is called a quantum-mechanical singlet state, their magnetic fields must be oriented in exactly opposite directions; most ordinary chemical bonds consist of such pairs of electrons.

When a lone electron is placed in a magnetic field, it is energetically favorable for the electron's magnetic field to be aligned in the same direction as the external field. If a sample of a given material contains a large number of lone electrons that do not interact with one another, then all those electrons will tend to align with any external magnetic field. This effect, which is known as paramagnetism, increases the strength of the total magnetic field within the sample.

The electride whose structure we have determined—that is, the electride made from cesium and 18-crown-6 crown ethers—behaves almost as if its electrons do not interact at all: it is strongly paramagnetic. This finding correlates well with previous evidence that its electrons are localized. Several other localized



MAGNETIC PROPERTIES of a material depend largely on interactions among electrons. Each electron spins and, like any spinning charged particle, generates a small magnetic field. Materials containing a large number of electrons that do not interact (*a*) tend to be paramagnetic: the fields of individual electrons tend to align with any externally imposed field, strengthening the total magnetic field within the material. In other materials (*b*) the electrons form pairs. The fields of the two electrons in a pair

© 1987 SCIENTIFIC AMERICAN, INC

electrides, such as one made from rubidium and the crown molecules designated 15-crown-5 (because the ring structure includes 15 atoms, five of which are oxygen atoms) and one made from potassium and the 15-crown-5 crown molecules, are also paramagnetic. In these compounds, then, the optical absorption spectra, electrical conductivities and magnetic properties all suggest that the electrons are localized at sites far enough apart to allow only weak interactions among electrons.

Long-Range Interactions

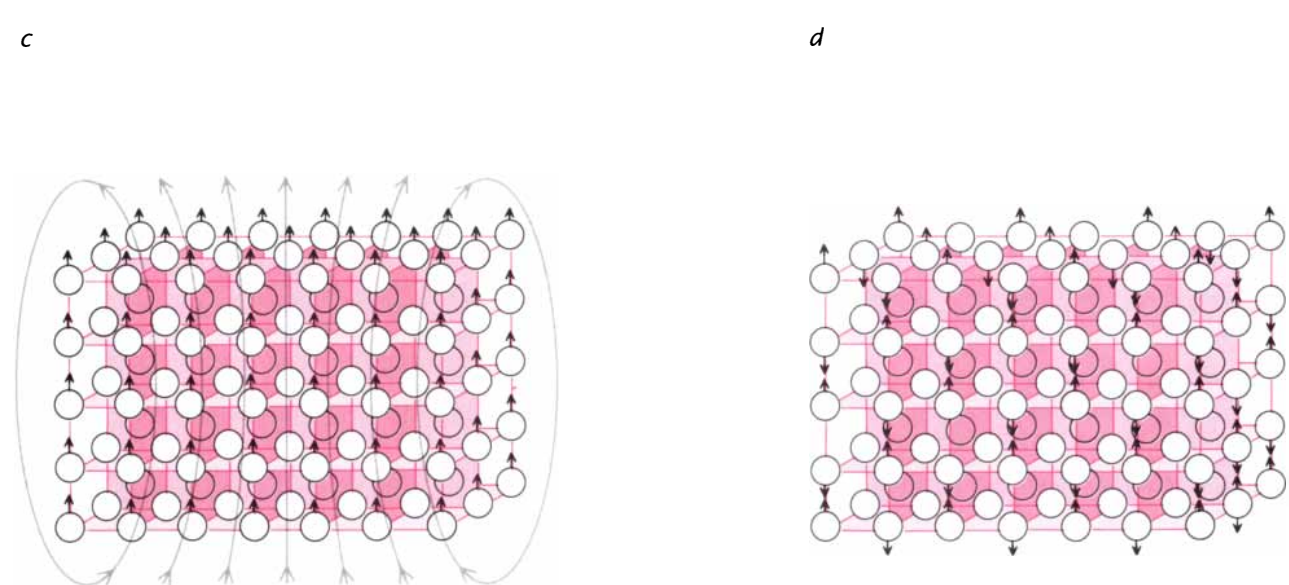
An interesting variation is seen in the compound made from cesium and the 15-crown-5 crown ethers, which seems, on the basis of optical and electrical properties, to have localized electrons. At temperatures above about 10 degrees Kelvin (degrees Celsius above absolute zero) the compound is paramagnetic, but when carefully annealed samples are cooled to 4.3 degrees, they display a more complicated effect called antiferromagnetism. Antiferromagnetism (like the related effect known as ferromagnetism) is the result of long-range interactions among many electrons in certain crystal lattices. When an antiferromagnetic material is cooled below a certain temperature, half of the electrons align spontaneously with their magnetic fields pointing in one direction and half align with their fields pointing in the opposite direction. Each electron's field points in the direction opposite to the fields of its nearest neighbors. (In ferromagnets, in contrast, the fields of all the electrons become aligned in the same direction.)

The appearance of such a long-range effect shows that there are small but significant interactions among anionic electrons in this compound. It is not completely clear why the related rubidium- and potassium-based electrides do not show the effect. Perhaps those compounds are simply more difficult to anneal into crystals that can support antiferromagnetism. It may be, however, that the more compact crown-ether sandwiches that can be formed around these smaller cations result in qualitatively different materials.

A different kind of magnetic behavior is found in the electride made from lithium and the cryptand designated [2,1,1] (because one of the chains that make up the "bars" of the crypt includes two oxygen atoms and the other chains include one). This compound is paramagnetic at high temperatures, but at lower temperatures the compound's paramagnetism decreases, approaching zero at zero degrees K.

Such magnetic behavior is seen in materials whose electrons interact in pairs. At high temperatures the electron pairs are broken up by thermal energy and the resulting lone electrons behave paramagnetically, but at low temperatures the electron pairs can form. The magnetic fields of the two electrons in a pair point in opposite directions and hence cancel each other. The pair as a whole therefore shows no paramagnetic behavior. The magnetic character of this electride suggests that the crystal has an underlying structure in which each anionic cavity is close enough to one other cavity for the electrons to interact as a pair but too far from other pairs to interact appreciably with them.

Finally, the electride made from potassium and the cryptand designated [2,2,2], which shows the optical and electrical behavior characteristic of a metal, also shows magnetic behavior one might expect of a metal or of a solid containing weakly bound electron pairs. Apart from certain other small effects, this electride exhibits a small paramagnetism that increases slightly with increasing temperature. In a system containing completely delocalized electrons one would expect weak paramagnetism that was independent of temperature. Although in such materials most of the electrons are paired at any temperature, some are unpaired and indepen-



point in opposite directions, canceling each other's effect on the total magnetic field within the sample. In some materials (*c*) large numbers of electrons interact over long ranges in such a way that they all become aligned, even in the absence of an external field. Such materials are said to be ferromagnetic. Another kind of long-range interaction is antiferromagnetism (*d*), in which the fields of half of the electrons point in one direction and the fields of the others point in the opposite direction.

© 1987 SCIENTIFIC AMERICAN, INC

dent, and they are the source of the paramagnetic effect. As the temperature goes up, the number of unpaired, paramagnetic electrons rises because some pairs are broken apart by thermal energy (because the electrons in metals are delocalized, pairs can be broken apart easily), but so does the likelihood that an unpaired electron will have enough thermal energy not to become aligned with an external magnetic field. The fact that the paramagnetism of this highly conductive electride is somewhat dependent on temperature suggests that if the electrons are delocalized, they may travel in pairs through the lattice.

Interactions with Nearby Nuclei

In understanding electrides one would like to know not only how the electrons interact with one another but also how they interact with the cations that are their immediate neighbors. How much time, for example, does a trapped electron spend in the vicinity of surrounding cations, away from the center of its anionic cavity?

One extremely powerful method for answering such questions is nuclear magnetic resonance. Atomic nuclei, like electrons, generate an intrinsic magnetic field. When a nucleus is placed in an external magnetic field, the energy of the nucleus as a whole is lower if its own magnetic field points in the same direction as the external field rather than in the opposite direction. Electromagnetic energy, however, can "flip" a nucleus from the low-energy state to the high-energy state. The nucleus will absorb only those photons that have the right amount of energy to induce the transition.

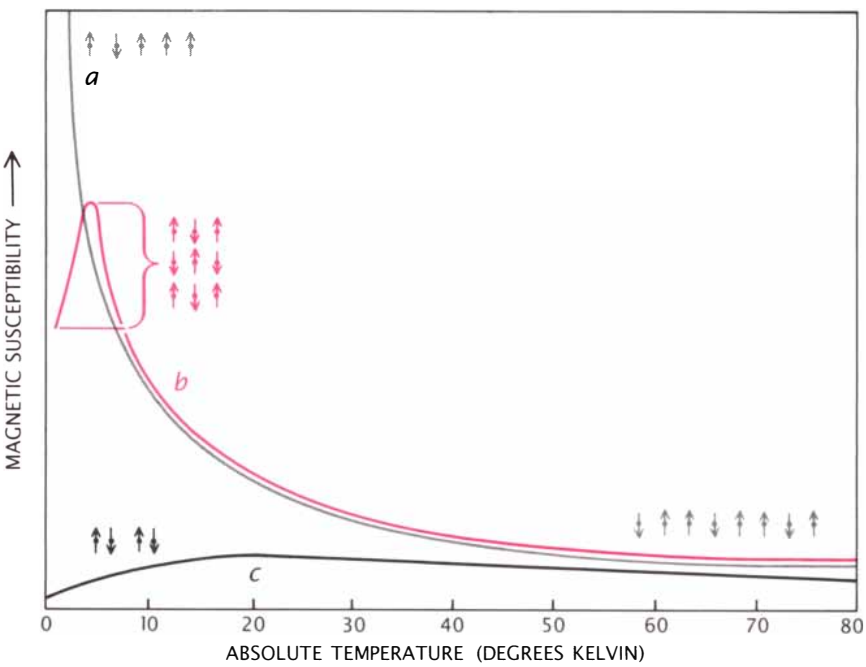
The energy of the transition depends on the characteristics of the nucleus, its environment and the strength of the magnetic field. By holding a material in a steady external magnetic field and determining the energy of the photons it absorbs, one can probe the characteristics of its constituent nuclei and their electronic environments.

By probing the electronic environment of the cations in an electride, nuclear magnetic resonance can reveal how much time the trapped electrons spend near a cation. Lone trapped electrons are paramagnetic—their magnetic fields become aligned with any external field—and so the magnetic field of such an electron near a nucleus will add to the external field to produce a greater total magnetic field in the vicinity of the nucleus. This strengthened field affects the amount of energy needed to flip the nucleus. The wavelength of the photons the nucleus will absorb is therefore a direct measure of the density of lone electrons in the nucleus's vicinity.

We compared the nuclear-magnetic-resonance spectrum of isolated cesium atoms with the spectrum of cesium cations in the electride of cesium and 18-crown-6 complexants and found that the trapped electron interacts very weakly with the nucleus. Each nucleus had a paramagnetic electron in its vicinity only about .03 percent of the time.

In a separate test we addressed another question: we tried to determine whether each cation interacts equally with all eight of the surrounding electrons, just as cations in true ionic salts interact equally with the surrounding anions. In answering this question we took advantage of the close structural similarity between the electride of cesium and 18-crown-6 complexants and the alkalide made from cesium, 18-crown-6 complexants and negatively charged sodium ions. We grew a sample of the cesium–18-crown-6 electride in the presence of sodium in such a way that some sodium anions substituted for trapped anionic electrons, and we analyzed the resulting crystals by nuclear magnetic resonance.

We found that some cesium cations were surrounded by only seven-eighths as much electron density as usual, other cations by only three-quarters as much and still others by only five-eighths as much. Our conclusion was that in these cases one, two or three of the eight anionic sites surrounding each cation were occupied by sodium anions rather than by electrons. Because removing any of the eight electrons had the same effect, this result showed that a cation



MAGNETIC SUSCEPTIBILITY of electrides (the degree to which an externally imposed magnetic field is reinforced by fields generated within the electride) gives information about the interactions of electrons in various electrides. The magnetic susceptibility of the electride made from cesium and 18-crown-6 (*a*) is very high at low temperatures, indicating that the material is paramagnetic, and so it can be deduced that this electride's trapped electrons are independent. The magnetic susceptibility of this electride falls at high temperatures as the magnetic fields of the electrons become thermally disordered. An electride made from cesium and a smaller crown ether (*b*) is paramagnetic at high temperatures, but it has the low magnetic susceptibility of antiferromagnetic materials at temperatures below 4.3 degrees Kelvin (degrees Celsius above absolute zero). The trapped electrons in this electride are close enough together to act as a large ensemble at low temperatures and show long-range interactions. In another electride, made from lithium and a certain cryptand (*c*), trapped electrons form pairs at low temperatures, lowering the magnetic susceptibility, but they do not display the properties expected of large ensembles. It can therefore be deduced that each electron is close to one other electron, but that each pair of electrons is far from other pairs.

© 1987 SCIENTIFIC AMERICAN, INC

in an electride does not bind particularly strongly to any one electron or vice versa.

These studies and others confirmed our theoretical picture: localized electrides are simple ionic salts in which all the anionic sites are occupied by trapped electrons. There seems to be no special relation between a trapped electron and any one of the surrounding cations, and the electrons spend the great majority of their time in and near the cavities rather than near surrounding cations.

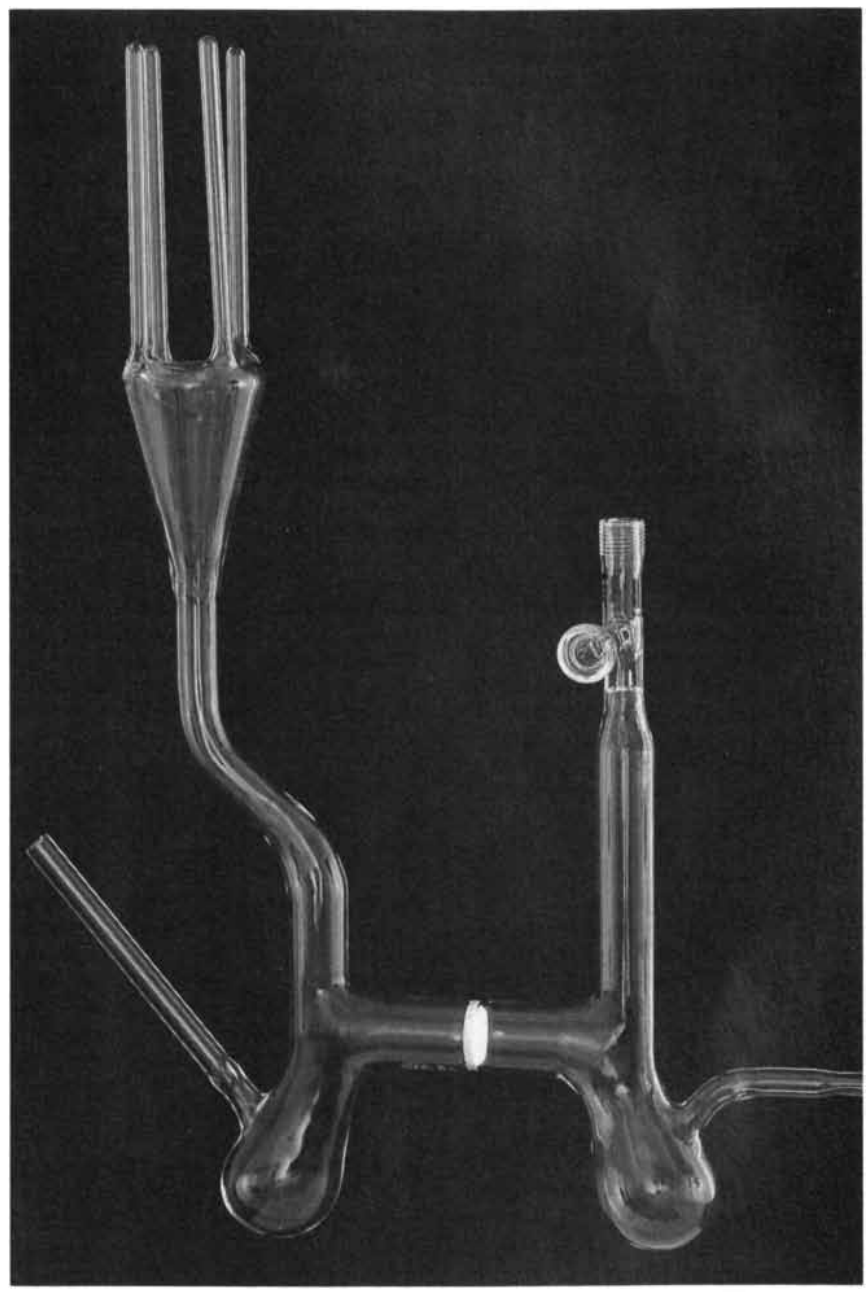
Possible Applications

In addition to their intrinsic scientific interest, electrides could someday be useful in a variety of applications. They can be dissolved in a number of solvents to yield solvated electrons (electrons that are free in solution). Solvated electrons in liquid ammonia are used extensively in industrial syntheses, but there are a number of reactions in which liquid ammonia cannot be used, either because it can sometimes donate protons or because one reactant is not soluble in liquid ammonia. The ability to dissolve electrides in other solvents could therefore be quite useful in a variety of syntheses.

Because the electrons in localized electrides are so weakly bound, these crystals could be effective as photosensitive detectors in infrared-sensing photomultipliers. In these devices an impinging photon liberates an electron, resulting in a small electric current. The same weak binding could also make electrides useful in solar-energy converters and as cathodes in batteries.

The main barrier to incorporating electrides into new electronic devices is their tendency to decompose. In our laboratory we are embarking on a major attempt to synthesize electrides that do not spontaneously decompose at room temperature. By their very nature all electrides (which contain loosely bound electrons) must react with air and moisture, which readily accept electrons. There is reason to believe, however, that we can produce electrides that are otherwise stable.

For example, we have done tests with a complexant called hexamethyl hexacyclen, which is similar to 18-crown-6 except that the oxygen atoms are replaced with nitrogen atoms that have auxiliary groups attached to them. We have found that an alkalide consisting of potassium cations, hexamethyl hexacyclen complexants and sodium anions is extremely stable: it remains so at temperatures as high as 45 degrees C. The next step will be to see whether we can synthesize stable electrides based on complexants of this type. If we do succeed in producing stable electrides, this new form of crystalline matter could have many useful technological applications, in addition to opening a new window on the basic understanding of interactions among electrons in solids.



SYNTHESIS APPARATUS makes it possible to grow relatively large electride crystals. The complexant (crown ether or cryptand) is placed in one bulb (*lower left*) and the metal that will provide the cations is placed in the other bulb (*lower right*). A solvent is introduced into the bulb holding the complexant, and the apparatus is tipped so that the dissolved complexant pours through a filter made of sintered glass onto the metal. The metal dissolves, and the apparatus is tilted again so that the entire solution pours back through the filter into the first bulb (the filter traps pieces of undissolved metal). When a weaker solvent is added and the solution is cooled, crystals of electride form spontaneously. The solvent is poured off, and the apparatus is tipped upside down so that the crystals will fall into thin glass tubes protruding from its top. The tubes are heat-sealed and stored individually. The glassware was made by master glassblower Keki Mistry.

© 1987 SCIENTIFIC AMERICAN, INC

Mimicry in Plants

There are flowers that look like insects and weeds that masquerade as crop plants. Mimicry in plants results from natural selection: it attracts pollinators or deters predators

by Spencer C. H. Barrett

Plant adaptations can be remarkably complex. Certain species of orchids, for instance, imitate female bees; other plants look and smell like dead animals, and still others have the appearance of stones. These strange adaptations to life represent just a few of the sophisticated means by which plants enhance their chances of survival.

The idea that one species can imitate another in order to deceive an enemy and thus escape predation has fascinated biologists for more than a century. This particular survival strategy was first described in 1862 by the English naturalist Henry W. Bates, who observed that certain brightly colored species of South American butterflies were almost identical in appearance but that some were poisonous whereas others were completely harmless. He surmised that the harmless species were masquerading as harmful ones in order to avoid being eaten and called the phenomenon mimicry; it is now known as Batesian mimicry.

Mimicry in plants has not been the focus of much attention until recently, partly because in the past plants were studied largely in relation to their physical environment and partly because mimicry is simply less common in plants than it is in animals. (Most likely the reason is that plants are sedentary and have a propensity to form clusters; consequently herbivores can learn the location of individual plants and discriminate between them.) Nevertheless, observations made in 1793 by the German naturalist Christian K. Sprengel provided an early clue that plants, like animals, imitate other species. Sprengel found that certain plants, particularly orchids, do not secrete nectar themselves but instead mimic the appearance of nectar-producing species living in the same habitat.

It was not until well into this century, however, when ecology emerged as a distinct discipline, that the relation of plants to other species in the environment became the object of detailed inquiry. Since then studies have revealed that plant-animal interactions run the gamut from associations that are mutually beneficial (as in the case of pollination systems) to associations that are antagonistic (as in the case of defense against herbivory). Mimicry can play a role in all these systems, and as botanists continue to study plant-animal interactions, particularly in tropical ecosystems, it has become clear that plant mimicry is far more widespread than had been supposed.

Mimicry, in plants or in animals, is a three-part system. There is a model: the animal, plant or substrate being imitated. There is a mimic: the organism that imitates the model. And there is a signal receiver or dupe: the animal that cannot effectively distinguish between the model and the mimic. Mimetic traits may include morphological structures, color patterns, behaviors or other attributes of the mimic that promote its resemblance to a model. That model may be either an unrelated species or an inanimate object, such as the background against which an organism spends most of its time.

Mimicry is not an active strategy on the part of an individual plant; flowers do not deliberately trick or deceive animals into visiting them. Mimicry arises as the result of evolution through natural selection and the occurrence of random mutations that lead over many generations to the appearance of favorable characteristics. If such genetically based traits help to camouflage a plant, for example, the plant is likely to have a survival advantage over other plants that are less well camouflaged. The plant will leave more descendants, thereby passing the advantage to the next generation. For natural selection to favor the evolution of mimicry, the mimic must derive a reproductive advantage from modeling itself after another organism or object; its fitness, measured as the number of offspring produced that survive into the next generation, must be increased as the result of deception.

This has rarely been documented in field studies of mimicry in animals, in part because keeping track of two or (more commonly) three interacting species can be logistically difficult. Plants, at least during the vegetative phase of their life cycle, are immobile and are therefore more easily observed and manipulated under experimental field conditions. In many cases of herbivory it is obvious that the mere survival of the plant automatically confers a fitness advantage over less fortunate individuals that are preyed upon. In pollination the number of pollinators attracted to a plant and the amount of fruit and seed produced can be considered estimates of reproductive fitness.

In most pollination systems both the plant and its pollinators benefit equally, a relationship known as mutualism. The plant offers its pollinators nutritional rewards in the form of nectar and pollen, and during feeding activities the pollinators transfer pollen from plant to plant within a species, leading to fertilization and the setting of seed. Certain plants exploit this relationship and are visited by pollinators in spite of the fact that they offer no nutritional rewards. The plants do so by mimicking any of a number of physical and chemical characteristics (including color, texture, scent and form) of nearby flowering-plant species that secrete copious quantities of nectar. The insects

© 1987 SCIENTIFIC AMERICAN, INC

are unable to distinguish between the model and the mimic and therefore pollinate both of them.

One example of such deception, whose basis is not immediately obvious to a human observer, was recently described by L. Anders Nilsson of the Institute of Systematic Botany in Uppsala. The model in this mimicry complex consists of several species of bellflowers in the genus *Campanula;* the mimic is the red helleborine orchid *Cephalanthera rubra.* Both flowers live together in dry, woody habitats throughout western Europe and the Mediterranean, where they are regularly visited by two species of solitary bees: *Chelostoma fuliginosum* and *C. campanularum.* Beyond this the two flowers bear little resemblance to each other. They differ in both shape and color (the orchid is rose-colored, the bellflower a violet-blue) and belong to separate plant families. Moreover, the bellflowers produce abundant nectar for the bees that visit them, whereas the orchids produce none.

Nilsson observed that male *Chelostoma* bees fly to the *Cephalanthera* flowers, alight and enter the flower headfirst, presumably searching for floral rewards. Not finding nectar, they back out of the flower and in doing so brush against packages of yellow pollen called pollinia. These become attached to the thorax of the bee by a sticky substance secreted by the flower's female sex organ, the stigma. When the bee flies away, it carries the pollinia with it and deposits its pollen on the stigma of the next orchid it visits. Nilsson tracked the bees and found that, by not discriminating between the two types of flowers, the bees successfully pollinate both species. He wondered why the orchids, which are distinctive to look at and lack nectar, were attractive to the bees that pollinate them.

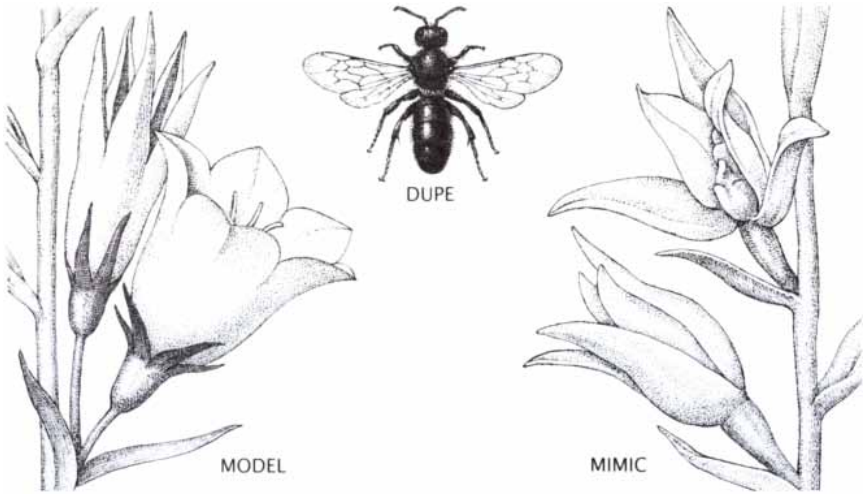
Nilsson solved the paradox with the aid of a spectrophotometer, a device that measures the wavelengths of light reflected from an object. Although the colors of the orchid and the bellflower look very different to human observers, he found that the reflectance curves for the two flowers are virtually identical within the

DECEIVED BY physical and chemical cues, the wasp *Campsoscolia ciliata* tries to mate with the orchid *Ophrys speculum.* The labellum, or lower lip, of the flower has roughly the size, shape and fuzziness of a female wasp; male wasps apparently cannot distinguish between them. As a result the orchid is pollinated by wasps that visit one flower after another searching for a mate.

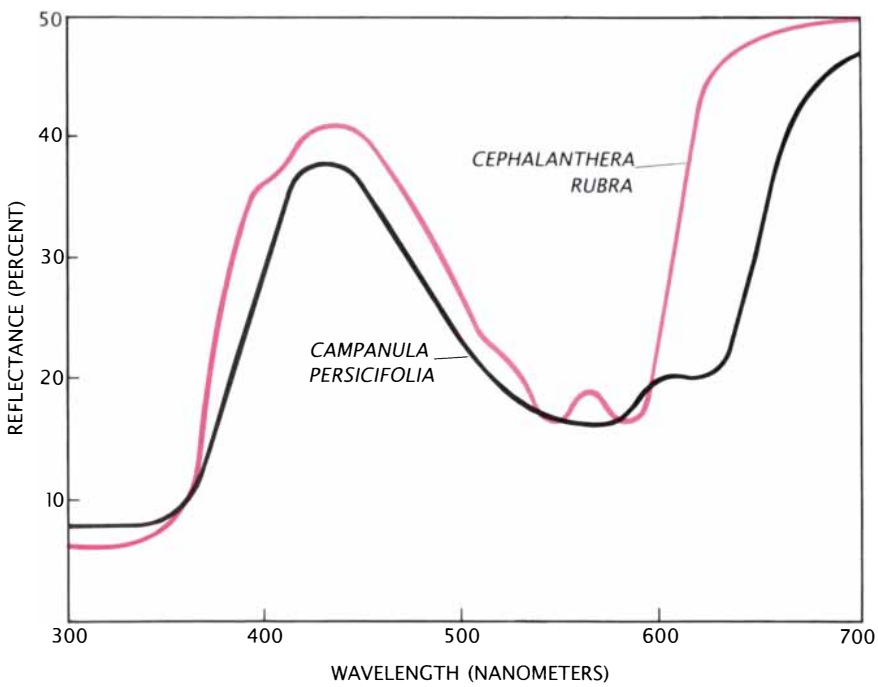
© 1987 SCIENTIFIC AMERICAN, INC

bees' visual range. The orchid appears red to the human eye because it reflects light at the red end of the spectrum (in the range of 600 to 650 nanometers), but bees see best at the ultraviolet end; like many insects, they do not distinguish among different red hues. To the bee, therefore, the orchid and the bellflower are identical; shape and scent are apparently of minor importance in this deception system.

Male *Chelostoma* bees are thus deceived into visiting *Cephalanthera* flowers, a behavior that is of little benefit to the bee but very beneficial to the plant: it is pollinated without having to provide floral rewards in return. Comparison of the reproductive success of the orchid in habitats where it coexists with bellflowers and their bee pollinators and where it does not shows that the orchids clearly increase their fitness by mimicking bellflowers. Fruit set in these areas is six times higher than it is in regions where the orchid exists without bellflowers.



MIMICRY IS A THREE-PART COMPLEX involving a model, a mimic and a dupe. Here the model is the bellflower *Campanula persicifolia* (*left*), which produces abundant nectar, a substance that is attractive to pollinators. The red helleborine orchid *Cephalanthera rubra* (*right*) produces no nectar, but it mimics the bellflower and thus deceives the dupe: the leaf-cutting bee *Chelostoma fuliginosum* (*center*), which visits it. The bee cannot distinguish between the colors of the two flowers and pollinates both species.



COLORS OF MODEL AND MIMIC are different to human beings, who can detect wavelengths of light at the red end of the visible spectrum (*right*), where the reflectance patterns of these flowers diverge. But at the ultraviolet-blue end of the spectrum (between approximately 300 and 550 nanometers), where bees see best, the reflectance curves of the two flowers are almost identical and the flower colors look the same to the bees.

Examples of plant mimicry considerably more bizarre than this occur elsewhere in the orchid family. In one type of deception the flowers of several orchid species imitate female insects. The orchids produce no nectar but instead emit a scent remarkably like the sex pheromone of the insect species they are mimicking. In addition the labellum, or lower lip, of the orchid flower tends to have a shape and a texture like those of the insect being imitated. In orchids that attract bees, for instance, the labellum is covered with hairs.

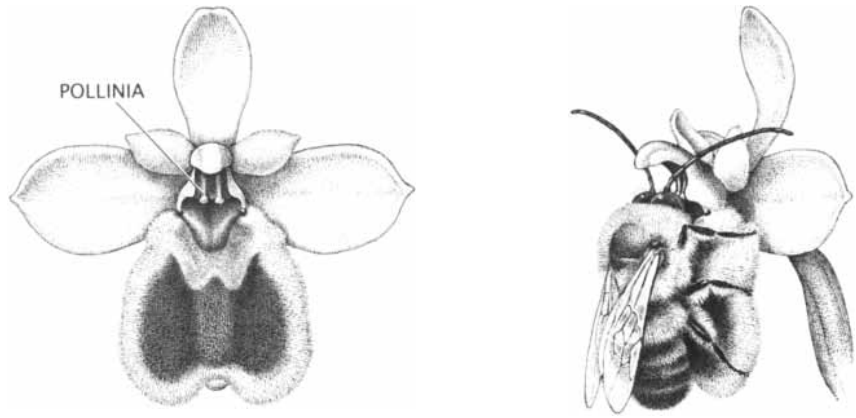
Several species in the European genus *Ophyrs* have evolved so closely with their insect pollinators that they produce a fragrance similar to, if not identical with, the sexual attractant of the female insect they are mimicking. In some cases individual orchid species are named after the species of insect (fly, wasp or bee) that pollinates them. Males, particularly if they have not mated with a female, are attracted to the orchid's flowers and attempt to copulate with them. From the plant's perspective, pseudocopulation is clearly an effective strategy. When the insect lands on the flower, it makes contact with the orchid's pollinia. Unsuccessful in the mating attempt, the insect flies off in search of a more appropriate mate, transporting the pollinia to another flower of the same species.

Pseudocopulation represents one of the most beguiling and complex forms of animal-mediated pollination in flowering plants. It has evolved independently on three separate landmasses (Australia, Eurasia and South America) and is particularly prevalent in the tropics, where many orchid species reproduce by this means. Why do insects, which receive no apparent benefit, participate in the arrangement? Should natural selection not favor the survival of males that are able to conserve their energy by distinguishing a flower from a female insect? One possible explanation is that flowering time for the orchid is closely synchronized with the emergence of adult males from the pupal stage. (Male insects often emerge before the females.) Emerging males compete heavily for

© 1987 SCIENTIFIC AMERICAN, INC

the few available adult females; under these circumstances their discriminatory powers are thought to be low. In a state of sexual frustration many males are readily deceived by the scent and appearance of the orchid flowers.

In stark contrast to the showy flowers associated with pseudocopulation is a pollination strategy in which plants assume some of the characteristics of rotting flesh. Such plants have evolved traits, including putrid odors, fleshlike colors and abundant hairs, that are repulsive to human beings but attractive to flies and other carrion-feeding insects. One of the most potent of these rotting-flesh mimics is *Amorphophallus titanum,* an eight-foot plant from Sumatra whose stench is so powerful that people are said to have fainted from sniffing it too closely.



PSEUDOCOPULATION between two Mediterranean species, the horned bee *Eucera nigrilabris* and the orchid *Ophrys tenthredinifera,* leads to pollination of the orchid. To a male bee the flower (*left*) looks like a female bee with her wings outspread. The bee grasps the flower's labellum and attempts to mate with it (*right*). While doing so he brushes against the pollinia: pollen sacs hanging from the upper lip of the flower that stick to his body. When the bee leaves, he carries the pollinia to another *Ophrys* flower.

Rotting-flesh mimics rarely offer nectar or other food rewards. Instead they deceive their pollinators by appearing to be a suitable food site for developing maggots, which normally feed on dead animal tissue. Female flies, fooled by the smell of rotting flesh, land on these flowers and in some cases even lay their eggs in them. In a manner reminiscent of the male insects that move from flower to flower seeking a mate, female flies go from one rotting-flesh mimic to another in search of egg-laying sites. As they progress from flower to flower they inadvertently collect pollen on their bodies and transport it to the flowers' stigmas, thus pollinating the species.

Some plants imitate other members of their own species, a form of deception known as Bakerian mimicry for Herbert G. Baker of the University of California at Berkeley, who first described it. For many years it has been known that members of the Caricaceae, or papaya family, are dioecious: they bear male and female flowers on separate plants. In the Caricaceae male and female flowers are strikingly different, a phenomenon known as sexual dimorphism. Male flowers in this family possess long floral tubes, whereas female flowers lack floral tubes and instead have free petals surrounding a large green ovary. Male flowers in the Caricaceae produce nectar but female flowers do not.

Baker wondered how the species was pollinated. What would attract an insect to the female flower, which produces no nectar or pollen? In studying the interactions of papaya plants with their pollinators he noted that the plants are pollinated mainly at twilight by insects such as moths in the family Sphingidae, which visit both male and female flowers equally. He concluded that to a sphingid moth flying above the plants when light levels are low the male and female flowers must look remarkably alike. Unable to distinguish between the two flower types, the moths are deceived into visiting the nectarless female flowers. Baker called the phenomenon "mistake pollination."

In most dioecious plant species male plants produce considerably more flowers than female plants. Kamaljit S. Bawa of the University of Massachusetts at Boston suggests that the skewed ratio of flowers in dioecious plants has resulted in intense natural selection for female flowers to conform to the search image developed by pollinators visiting male flowers. This type of deception is consistent with the rules of mimicry: the model (the male flower) in the family Caricaceae is much more abundant than its mimic (the female flower). For that reason pollination biologists believe mimicry in dioecious plants may be more widespread than is supposed, and it is interesting that male and female flowers in animal-pollinated plants tend to look alike, whereas those that are wind pollinated often do not.

Not all plant mimicry is one-sided as in Batesian mimicry, where one species obtains the advantage by imitating a second species. Another type of mimicry, observed commonly in animals but rarely in plants, is known as Müllerian mimicry (after Fritz Müller, the 19th-century German zoologist who first described it). In this type of mimicry a number of unrelated species resemble one another in appearance and behavior. By forming a group of look-alikes the plants are able to take advantage of a common "advertising style," a strategy that attracts more pollinators than would be the case if each plant advertised alone.

Müllerian mimicry differs from Batesian mimicry in that there is no deception, no clear distinction between model and mimic, and all participants contribute equally; that is, all offer floral rewards of comparable value. For this reason some botanists maintain that Müllerian mimicry is not really mimicry at all but simply a function of the fact that organisms living in the same habitat evolve in similar ways, a phenomenon known as convergent evolution. (A well-known example of convergence is the evolution of succulence in desert plants. In response to the need to conserve water in arid environments, various unrelated plant species have evolved thick leaves and fleshy tissues.)

There are also floral associations that appear to combine elements of both Batesian and Müllerian mimicry systems. One example that falls into this category but is still untested is a complex of three orange-yellow butterfly-pollinated plants: *Lantana camara* in the verbena family, *Asclepias curassavica* in the milkweed family and *Epidendrum radicans* in the orchid family. The three plants are commonly found together in road-

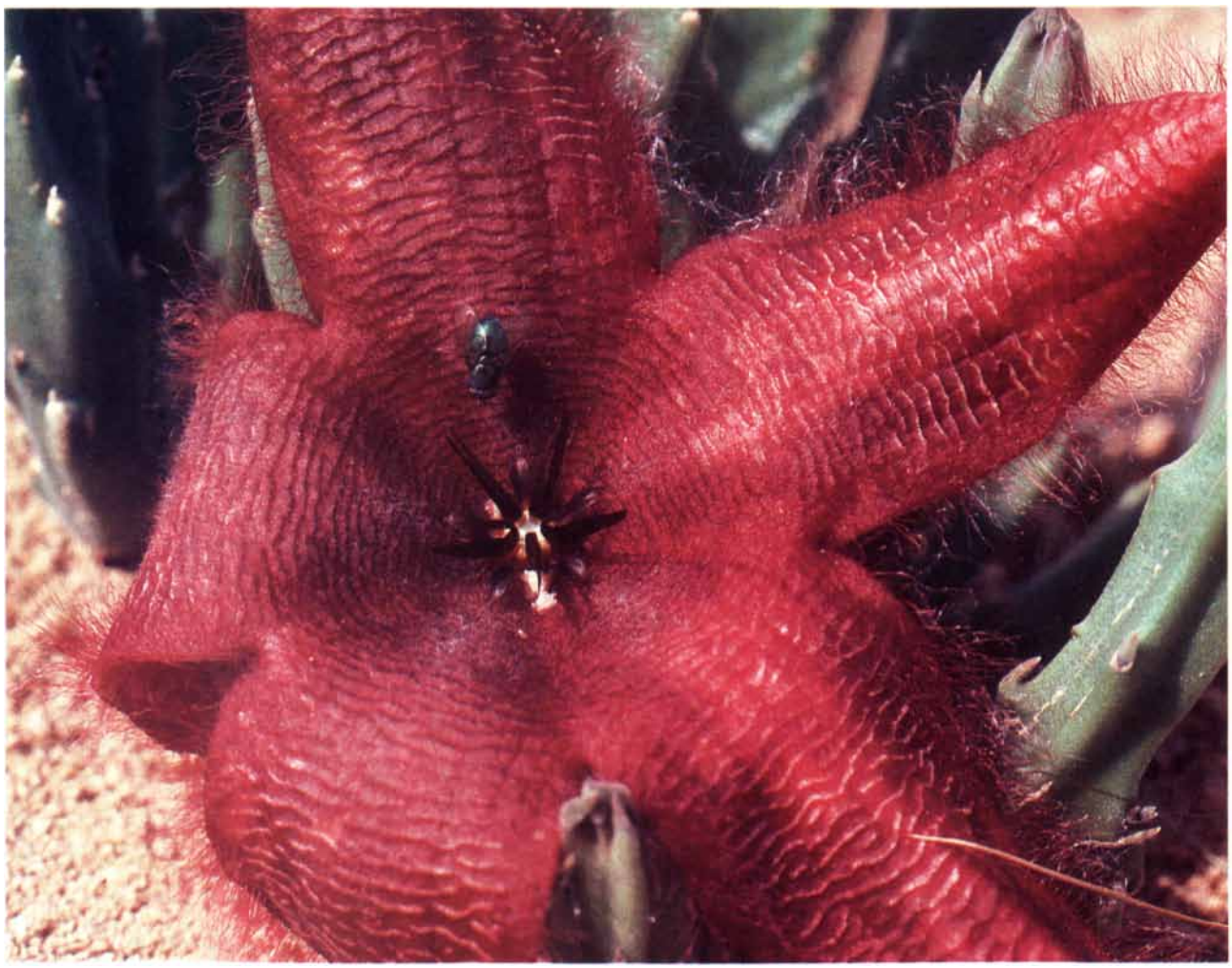
© 1987 SCIENTIFIC AMERICAN, INC

side communities in the American tropics, where they are visited by the same individual butterflies.

Lantana and *Asclepias,* which produce abundant nectar, are thought to be Müllerian mimics of each other. *Epidendrum,* which produces no nectar, is considered a Batesian mimic of the other two. The three species meet the outward criteria of a mimicry complex: they inhabit the same area and have flowers that are similar in size and color. Nevertheless, recent work has failed to demonstrate that each species is pollinated more (and therefore produces more offspring) when it occurs in association with the other two species than when it occurs in isolation.

Plants do not always mimic other organisms and need not always deceive for the purpose of pollination. In some species mimicry has evolved in response to predation by animals. Certain plant species resemble the shape, size, color and even the texture of an inanimate background to such an extent that they are camouflaged against potential herbivores. For example, several hundred species in the southern African ice-plant family, the Mesembryanthemaceae, have evolved a striking resemblance to stones and pebbles. Stone mimicry in the family is most developed in the genus *Lithops:* the entire plant, not just part of it, resembles a stone. *Lithops* plants are virtually impossible to spot against low-growing vegetation, particularly during the dry season, when they shrivel and often become lightly covered with sand.

Delbert Wiens of the University of Utah suggests that animals such as ground-foraging birds, small mammals or ungulates that once roamed the plains of Africa in large numbers may have provided the selective force that has led to the radiation of these mimetic forms. Grazing on low-growing vegetation, the animals might have devoured anything remotely plantlike in appearance, particularly during the dry season, when many plants wilt and die. This hypothesis is supported by the fact that *Lithops* flowers (which are readily visible) are short-lived and appear only during the wet season, when there is abundant herbage for grazing animals. The difficulty of identifying the selection pressures responsible for stone mimicry is a reminder that life is continuously evolving. When the intricate relation between model, mimic and dupe is disrupted, for instance by human disturbance,



CARRIONLIKE PLANT ***Stapelia nobilis*** from southern Africa has the color and stench of rotting flesh. To an insect the flower's red petals, which are covered with fine hairs, look and feel like fur. The female blowfly seen here, apparently unable to tell the difference between the plant and an animal carcass, has just laid a mass of white eggs at the bottom of the flower. The maggots that hatch from these eggs need animal protein for development and will die soon after birth for lack of a suitable food supply.

© 1987 SCIENTIFIC AMERICAN, INC

only parts of the original mimicry complex may remain intact. In these cases the striking resemblance between models and their mimics is an intriguing glimpse of past evolutionary processes that are no longer in operation today.

Not all mimicry among plants is a reflection of evolutionary events in the distant past. One of the most destructive predators of plants is the species *Homo sapiens,* and several studies of cultivated crops indicate that farming practices have had a powerful influence on the genetics and evolution of certain species of weeds. Coincident with cultivation came the need to control weeds that invade agricultural land and reduce crop yields by competing with crops for nutrients and light.

Since plants were first domesticated by human beings, farmers have removed troublesome weeds from their crops by hand, and in spite of the widespread use of herbicides, hand weeding is still practiced, particularly in less developed countries. In the process of removing weeds, whether by hand or by hoe, agricultural workers are faced with the problem of distinguishing between the crop and the weed. In the majority of cases it is easy for the agricultural worker to distinguish between them. When the crop and the weed species are similar in appearance, however, the weed may be overlooked through mistaken identity with the crop. This is most likely to occur when the individual weeds are scattered throughout the crop at relatively low density, but it may also occur at high weed densities when the toil of hand weeding lowers workers' discriminatory powers, leading to more mistakes. In either case, weeds that more closely resemble the crop have a better chance of survival.

My own interest in this absorbing subject came about while I was studying the problem of wild-rice infestations in cultivated rice fields in Swaziland. During the 1950's two species of wild rice, *Oryza rufipogon* and *Oryza punctata,* had invaded the rice fields, where they multiplied and spread rapidly. Because the wild species are similar in appearance to cultivated rice, the full extent of their invasion was not recognized until the weed populations were too large to be effectively controlled. My observations of hand weeding by agricultural workers in the fields indicated that distinguishing between the different types of rice plants was very difficult and often the weed rices were left standing.

STONE MIMICRY is best developed in the genus *Lithops* from southern Africa. From a distance these three plants are difficult to distinguish from the pebbles that surround them; during the dry season they shrivel and are almost invisible even at close range.

Weeding in rice paddies, as in most cultivated fields, is usually undertaken during the early stages of crop growth, when competition between the rice crop and invading weeds is greatest. Therefore most weed mimics imitate the seedling stage in the plant's life cycle and are indistinguishable from the crop until the flowering stage, by which time most of the damage has been done.

In Swaziland I observed that under certain conditions cultivated rice and the wild rice *O. rufipogon* hybridize and produce viable offspring. This makes the job of distinguishing between crop and weed particularly troublesome. The complexities of hybridization between cultivated and weed rices have been extensively studied by scientists at the National Institute of Genetics in Japan. They found that certain cultivated traits—such as flowering at the same time as cultivated rice, synchronous germination and reduced shattering ability (the propensity of the rice seeds to fall prematurely from the head of the rice plant)—were expressed in crop-weed hybrids. Unfortunately the hybrids retain the traits that make wild rice inedible: their grains are small and have extremely hard husks that are difficult to mill. Even more troublesome is the fact that domesticated traits expressed in the hybrids allow them to persist and proliferate, making the weedy rices a still greater agricultural problem.

In an attempt to eradicate hybrid weeds from rice fields, plant geneticists in India developed purple-leaved strains of cultivated rice. They theorized that this would be a simple and effective means by which workers weeding rice fields by hand could differentiate the weed species from the edible species. Unfortunately the natural process of hybridization between the crop and the weed, in association with hand weeding, soon led to the appearance of purple-leaved weedy rices and color was rendered ineffective as a means of distinguishing between the different species.

Another serious pest of cultivated rice is barnyard grass. Barnyard grasses (*Echinochloa*) are among the world's most widespread and noxious weeds, infesting a wide variety of crops including corn, cotton and fruit trees as well as rice. They can usually be recognized by their reddish leaf bases, drooping leaves and

© 1987 SCIENTIFIC AMERICAN, INC

early flowering time. Various mutations resulting in the loss of these characteristics, however, have arisen among species of barnyard grass that infest Asian rice fields, making them so similar to cultivated rice that they are almost indistinguishable from it, particularly during the vegetative growth period. It is possible to differentiate between them, but only by examining cultivated rice plants closely for the presence of a ligule, a short appendage at the base of its leaves. This requires more time than is generally available to hand weeders, and so the weed species are often overlooked.

Today two of these rice mimics, *Echinochloa oryzoides* and *E. phyllopogon* (once called *E. oryzicola*), are obligate rice-field weeds that have spread from Asia to many of the major rice-growing regions of the world, including California. It appears that they entered the U.S. as contaminants of rice seed not long after the beginning of rice culture in the state, because they appeared as weeds in early rice trials at the Biggs Rice Research Station in California's Central Valley. Today they are one of the most serious weeds of rice in California. Although hand weeding is no longer practiced in that state, the close resemblance of barnyard grass to rice may have indirectly furthered its spread through California. Rice seed from fields thought to be weed-free but infested with barnyard grass was certified for distribution around the state in the early 1900's.

From California the mimic *E. oryzoides* is thought to have been exported as a rice-seed contaminant to Australia, where it was first collected at the rice reserach institution at Leeton in New South Wales in 1938. Today it is found scattered throughout the rice-growing areas of that state and has the potential of becoming a serious agricultural problem. Anthony Brown of the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia and I have examined the genetic variation of several enzymes present in populations of *E. oryzoides* from California and Australia. Our data indicate that populations in Australia represent a highly limited extract of the genetic information present in California populations of the same species. We concluded, therefore, that the variety of *E. oryzoides* found in Australia could have come only from ancestral stock in California.

THREE-WEEK-OLD SEEDLINGS of cultivated rice (*left*) are compared with two barnyard grasses: a mimic, *Echinochloa phyllopogon* (*middle*), and a nonmimic, *E. crus-galli* (*right*). Although the barnyard grasses are closely related, *E. phyllopogon* resembles rice more than it does *E. crus-galli*. Distinguishing between *E. phyllopogon* and rice is tricky: rice has a ligule (a small appendage at the base of its leaves) that *E. phyllopogon* lacks. *E. crus-galli*, on the other hand, is easily recognized by its reddish purple stem.

Mimicry is not limited to the vegetative stage of a cultivated crop; it can also be expressed at several other stages in a plant's life cycle. For example, in the intensive small-scale farming that characterizes the growing of cultivated rice in many parts of the world, particularly Asia, the transplantation of seedlings by hand from nursery beds to rice paddies provides an additional opportunity for human selection to occur.

Mimicry may also develop at the seed stage. Seed mimicry was first described by Russian botanists in the early 1900's based on observations of a variant of the weed *Camelina sativa* (var. *linicola*) found in flax fields in the Soviet Union. *Camelina sativa* plants have thin, unbranched stems, narrow, pale leaves and look like flax plants. The resemblance, however, is not the result of selection by hand weeding, which is not practiced in flax fields, but simply adjustment to life among the densely packed and shaded stands of flax. Mimicry in *C. sativa* involves the seeds of the weed, which resemble those of flax in several important respects.

Camelina seeds ripen at the same time as those of the flax plant, and because the fruits in which they occur are nondehiscent (they remain attached to the plant), they are harvested along with the crop. Normally seeds other than those of flax are removed during the processes of threshing and winnowing, but *Came-*

© 1987 SCIENTIFIC AMERICAN, INC

HAND WEEDING OF RICE is still practiced in many less developed countries by workers who spend long hours pulling weeds from rice paddies. Rice-mimicking weeds, overlooked under the combined pressures of time and fatigue, have become serious agricultural problems in some areas. The women in this photograph, made in 1969, are hand weeding rice fields in Swaziland. By the mid-1970's, rice cultivation in Swaziland ceased entirely because of the weed problem posed by the wild rice species.

lina seeds have evolved winnowing properties similar to those of flax. The seeds of the two species look different but the weed and flax seeds are blown the same distance by the winnowing machine. As a result the seeds are mixed together and sown the following season, ensuring that the mimicry complex continues. In various parts of the Soviet Union the seeds of different flax varieties have different characteristics. Remarkably, the *Camelina*-seed mimics in each region have evolved in parallel with their models and so persist as weeds.

With the introduction of more sophisticated winnowing machines as well as improved varieties of crops, it is likely that the intimate relationship between crops and their weeds will break down—at least temporarily. For example, the introduction of early-maturing dwarf varieties of cultivated rice in California fields makes it easier to recognize contamination by *E. phyllopogon*. The development of rice varieties, which bear seed before the seeds of its mimic are mature, are likely to lead to loss of the weed mimic altogether.

Interaction between species is always unpredictable and a change in one species frequently elicits a change in the other. Although it is possible the new selection pressures acting on *E. phyllopogon* will lead to the evolution of mimics that match their flowering time with that of rice, my own studies suggest this is unlikely: populations of *E. phyllopogon* in California contain relatively little genetic variation for such traits as development time, and this limits their capacity for evolutionary change.

Mimicry in agricultural systems is not always detrimental to human beings. Some weed mimics have, by virtue of their similarity to crop species, become crops of major importance themselves. N. I. Vavilov of the Academy of Sciences of the U.S.S.R. was the first to observe in the late 1920's that in early cereal crops such as wheat, weeds resembling rye and oats were harvested and accidentally exposed to selection by man for domesticated traits including large seeds, rigid panicles and nonshattering habit. With time the weeds gradually evolved into useful crops in their own right and today are entirely dependent on man for their survival. Vavilov used the term "secondary crop" to designate those domesticated plant species that were originally derived from mimetic weeds, and the entire process, which has no real counterpart among mimetic phenomena in more natural ecosystems, is referred to as Vavilovian mimicry. That evolution such as this can occur so quickly provides compelling evidence for the evolution of adaptation by natural selection.

Mimicry in plants involves many types of ecological interactions and covers a broad gamut from pollination strategies based on deception, to the avoidance of predation by camouflage to the survival of weeds based on selection by human beings for plants that mimic crop species. In each case mimicry results from the interplay between random genetic mutations and the selective pressures of the environment in which an organism lives. Thus the occurrence of mimicry in plants is a vivid demonstration of the power of natural selection as a guiding force in evolutionary change.

© 1987 SCIENTIFIC AMERICAN, INC

No other system of keeping up can compare with SCIENTIFIC AMERICAN *Medicine.*

YOUR SYSTEM: a time-consuming, futile struggle to keep up with the information explosion.

The classic texts are convenient references—but the information they contain is obsolete before publication.

Like many physicians, you probably rely on the texts you first used in medical school. But even using the most recent editions, you find material that no longer reflects current clinical thinking—and they lack the latest information on such topics as herpes, oncogenes, AIDS, and photon imaging.

Reading stacks of journals alerts you to recent developments—but can't give you quick answers on patient management.

Struggling through the hundreds of journal pages published each month—even on only the really significant advances in the field—is arduous and memory-taxing. And it's a task that costs physicians valuable time—their most precious resource.

Review courses cover clinical advances—but, months later, do you recall the details of a new procedure or unfamiliar drug?

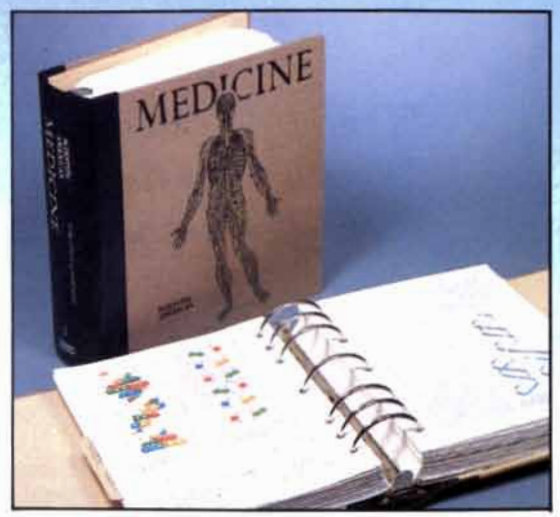
Seminars can also be costly and make you lose valuable time away from your practice—expenses that may amount to several thousand dollars. And, the speaker's skill often determines how much you learn.

SCIENTIFIC AMERICAN

MEDI

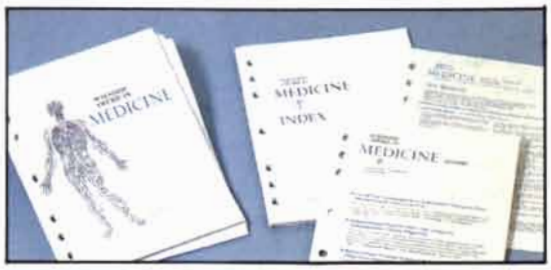
© 1987 SCIENTIFIC AMERICAN, INC

OUR SYSTEM: a rewarding, efficient way to keep yourself up-to-date—and save hundreds of hours of your time for patient care.



A comprehensive, 2,300-page text in two loose-leaf volumes, incorporating the latest advances in medical practice as of the month you subscribe.

This superbly designed, heavily illustrated resource, called "the best written of all [the internal medicine] books" by *JAMA* (251:807, 1984), provides a practical, comprehensive description of patient care in 15 subspecialties. And, because the text is updated each month, the clinical recommendations reflect all the current findings. A practice-oriented index and bibliography of recent articles further enhance the efficiency of the text.



Each month, six to nine replacement chapters to update your text *plus* new references, a four- to five-page news bulletin, and a completely new index.

You'd have to read hundreds of journal pages each month—and memorize the contents—to get the same information SCIENTIFIC AMERICAN *Medicine* contains. With our updated text, you read only the information you really need. Our authors, largely from Harvard and Stanford, sort through the literature and monitor developments, incorporating the significant advances into our chapters.

At no additional cost, a 32-credit CME program, to save you valuable patient-care time and the expense of attending review courses.

Earn 32 Category 1 or prescribed credits per year with our convenient self-study patient management problems; each simulates a real-life clinical situation. Choose either the complimentary printed version or, at a modest extra charge, the floppy-disk version for your IBM® PC/PC Jr., Macintosh™, or Apple® IIe or II+ (or compatible).

Your subscription includes:
the two-volume, 2,300-page loose-leaf text and, each month, six to nine replacement chapters, a newsletter, new references, and a completely revised index.

Or call toll-free: 1-800-345-8112
(in Penn. 1-800-662-2444)

CINE

SCIENTIFIC AMERICAN MEDICINE
415 MADISON AVENUE, NEW YORK, N.Y. 10017

☐ **Yes,** I'd like to try the SCIENTIFIC AMERICAN *Medicine* system. Please enter my subscription at a first-year price of US$245*.

☐ Also enroll me in the CME program. I'd prefer:
- ☐ the printed version at **no additional charge.**
- ☐ the floppy-disk version at an additional $US97.*

Computer type or compatible:
- ☐ IBM® PC/PC Jr. (128K RAM, DOS 2.0/2.1)
- ☐ Macintosh™ (512K RAM)
- ☐ Apple® IIe (64K RAM)
- Apple® II+ (64K RAM) 80-col. card by: ☐ Apple® ☐ Videx®

☐ Enroll me in the floppy-disk CME **only** at US$152.* (Note model above.)

☐ Check enclosed* ☐ Bill me

Signature _____

☐ VISA ☐ MasterCard Exp. Date _____

Account No. _____

Name _____

Specialty _____

Address _____

City _____ State _____ Zip Code _____

*Add sales tax for Calif., Ill., Mich., or N.Y. Allow 8 weeks for delivery. Add US$10 for shipping to Canada. IBM is a registered trademark of International Business Machines Corporation. Apple is a registered trademark of Apple Computer, Inc. Videx is a registered trademark of Videx, Inc. Macintosh™ is a trademark of McIntosh Laboratory, Inc. and is used by Apple Computer, Inc. with its express permission.

7L

© 1987 SCIENTIFIC AMERICAN, INC

The Swahili Corridor

In the 10th century Swahili sailors of the East African coast forged a route that brought gold, ivory and quartz to Europe. Their efforts contributed to the flowering of the Middle Ages

by Mark Horton

For most people the term renaissance calls to mind images of 15th-century Italy and names such as Leonardo da Vinci and Michelangelo Buonarroti. Five centuries earlier, however, on the shores of the Mediterranean there had been a cultural flowering that also deserves to be called a rebirth. In the 10th century the Holy Roman Empire was founded in southern Germany, the Byzantine empire was at its height and the Fatimid rulers of North Africa extended their control to Cairo. All these political developments were accompanied by great cultural revivals. Arts, crafts and architecture rose to new levels of complexity and subtlety. Such developments called for new and exotic materials, many of which were not available on the shores of the Mediterranean. Where were they to come from?

For at least three materials—gold, ivory and rock crystal (transparent quartz)—the answer was Africa. To bring them from there to the Mediterranean required a vast international trading network that reached from southern Africa to Sicily. One of the key elements of this network, whose existence has only recently been demonstrated, was provided by Swahili traders. They obtained gold, ivory and rock crystal from their inland contacts and carried them north to the Muslim merchants who relayed them to the Mediterranean world. The Swahili, an African people, were converted to Islam by their trading partners. Partly as a result of their conversion, they emerged as a cohesive social group that anchored the new trade route and helped to lay the groundwork for the 10th-century renaissance in Europe.

During that renaissance (which amounted to the beginning of the Middle Ages) gold, ivory and rock crystal appeared in the Mediterranean on a scale unparalleled since Greek and Roman times. Of the three, the history of ivory is the best-documented. Ivory had been widely used in the ancient world, and there are many classical and early Byzantine ivory carvings dating from the first century A.D. to as late as A.D. 600 Soon after 600 the supply from North Africa and the Red Sea seems to have dried up; only a handful of genuine elephant-ivory pieces are known in Europe from the three centuries following 600. Ivory craftsmen continued their art using substitutes such as bone, walrus ivory, narwhal tusks and even fossil mammoth ivory from Siberia.

In about 960 a radical change took place. Within a single decade Europe was flooded with magnificent examples of the carving of genuine elephant ivory. The revival appeared both in the Christian world, including Byzantium and the court of the Holy Roman Emperor, and in the Islamic world, particularly Spain and Sicily. Such a sweeping change can best be explained as the appearance of ivory from a previously untapped source. A clue to what the source might have been is the fact that many of the new pieces are more than 110 millimeters across, a diameter achieved only by the tusks of the African elephant. Indeed, documentary evidence from contemporary Muslim travelers indicates that much of the ivory came from East Africa.

Rock crystal shows a similar pattern. Like elephant tusks, clear crystal was a material that had been carved and polished since classical times. The surviving pieces are small and the crystal, probably from European sources, is often cloudy and imperfect. Late in the 10th century the crystal-carving industry was transformed. From the workshops of Fatimid Egypt came an outpouring of the finest rock-crystal working on a lavish scale. The earliest examples stem from about 975, and over the next 60 years production was so great that more than 100 surviving objects can be attributed to these workshops. The features of these fine pieces are their great size and the clarity and perfection of the crystal. As in the case of ivory, contemporary accounts suggest that the newly available crystal came from East Africa.

Although ivory and crystal were highly valued, gold was the most important product for the Mediterranean economy. It was both a medium of exchange and a medium of art. As a trade item it had great influence, because its possession could attract merchants and their goods. In that respect the Fatimid gold coins, minted throughout North Africa, were particularly potent. The reason was that the Fatimid metallurgists, working under strict control of the rulers, were able to produce a dinar (the unit of currency in the Islamic world) that was 96 percent pure. Such purity, far greater than that of any rival coinage, drew commerce in the southern Mediterranean to the Muslim ports rather than to the Christian ones.

Yet to draw commerce the operators of the Muslim ports had to surrender their dinars in exchange for goods, and so the supply of gold required continual replenishment. The traditional view has been that most of the gold for the Fatimid mints came from West Africa. Without question West Africa was an important source, but a variety of evidence suggests that East Africa was equally important. The Yemeni historian Al-

© 1987 SCIENTIFIC AMERICAN, INC

Hamdani, writing in 942, states that the world's most productive gold mine is in Ghana, in West Africa. But Al-Hamdani also makes statements that must be interpreted as referring to an extensive, secret sea trade carrying gold from sources around the Indian Ocean.

The array of evidence from art history, documents of the time and other sources makes it clear that East Africa was the new source of rock crystal, ivory and gold for the 10th-century Mediterranean world. Curiously, although the evidence seems unimpeachable, the connection between East Africa and the Mediterranean has been recognized only in the past several years. In that brief period archaeological discoveries by me and by others have confirmed the route's existence and demonstrated the central role of the Swahili in establishing it.

The Swahili ports and settlements extended 3,000 kilometers along the East African coastline. Modern archaeological surveys have revealed more than 400 sites along this thin line of coastal settlement that were occupied well before the Portuguese arrived in 1498. Many of the sites were small village communities. Several, however, were highly urbanized, with hundreds of stone houses and populations greater than 10,000. These were the commercial centers where the oceangoing dhows from the Indian Ocean made their landfall.

This coastal aggregation, which I have called the Swahili Corridor, was established by the ancestors of the Swahili who still live in the region. They were a maritime people who built boats of their own design and sailed them along the coast, following the seasonal monsoons and currents. Communication by sea along the coast resulted in the establishment of a common culture in the Swahili Corridor. An important part of that culture was Islam. According to contemporary Arabic descriptions, the Swahili traders were Muslims who followed a way of life comparable to that seen in the Middle East.

Where did these Muslim traders come from? Their origins have always been controversial. Local African traditions assert that the towns were founded by refugees fleeing upheavals in the Muslim Middle East in the eighth century. Some modern historians have argued that the Swahili were Arab colonists who intermarried with the Bantu farmers of the coast. My excavations at Shanga on Pate Island off the coast of Kenya, however, show that the "Arab" theory is incorrect: the Swahili were African in origin.

The work at Shanga occupied five digging seasons beginning with that of 1980. The site was chosen partly because it had a complete set of 14th-century ruins on the surface. Test pits in the center of the site revealed

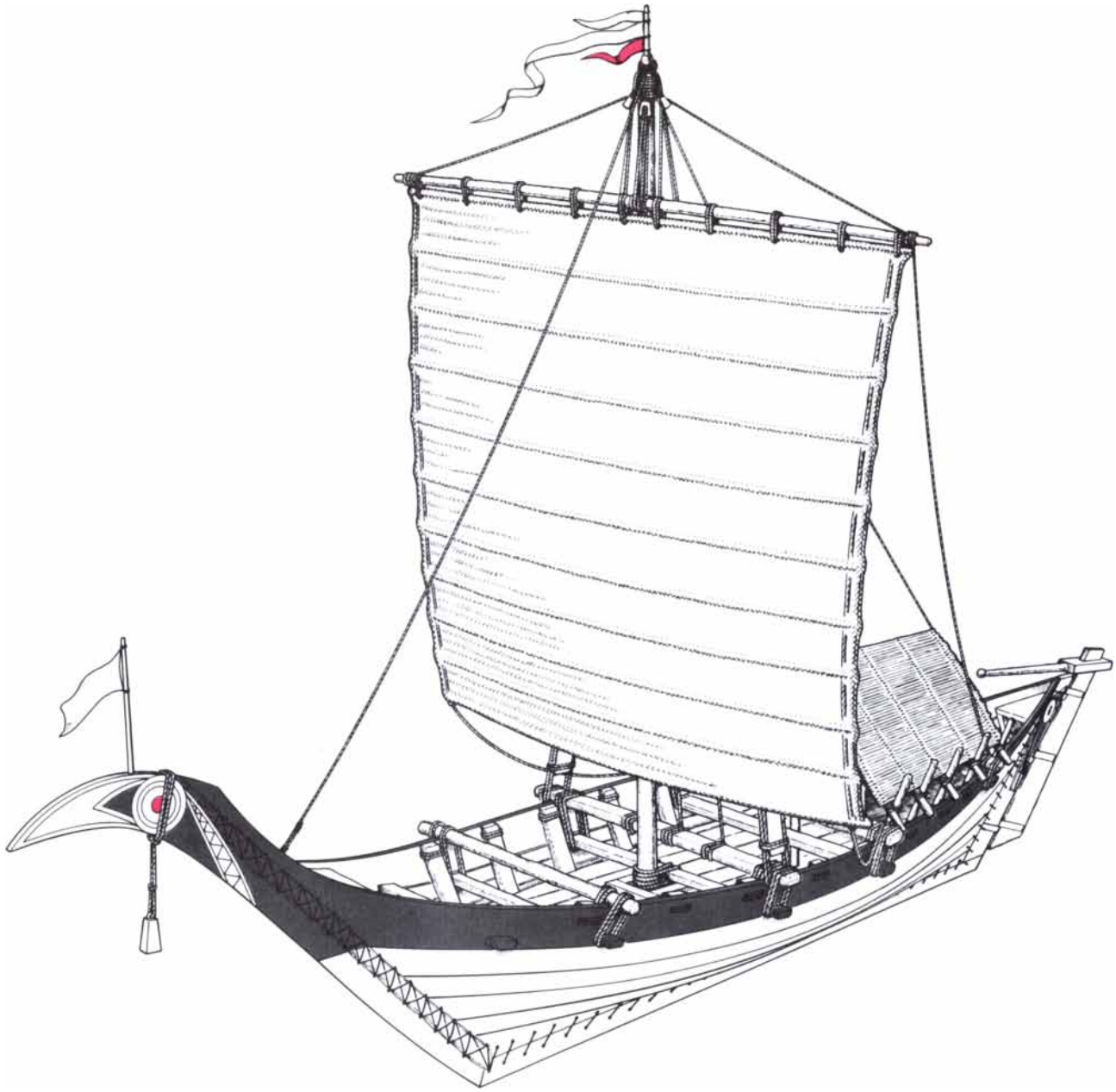
IVORY CASKET is an early example of a 10th-century revival of ivory carving in the Mediterranean world that may well have been due to new sources of ivory from Africa. The silver-mounted box, 4.5 centimeters high, was carved from a single piece of ivory. Probably intended to hold jewels, it was made in the royal workshops at Madinat al-Zahra in Spain for the daughter of the Umayyad ruler Abd al-Rahman III. The decorations include an inscription in the Arabic script called Kufic. The inscription indicates that the casket was made after the death of Abd al-Rahman, which took place in 961. The ivory for such beautiful objects was probably brought to the Mediterranean by an international trade in which the Swahili merchants of East Africa had a key role.

© 1987 SCIENTIFIC AMERICAN, INC

five meters of deposits underneath the ruins. It turned out that the earliest levels dated from the eighth century; 25 distinct occupation phases were recorded.

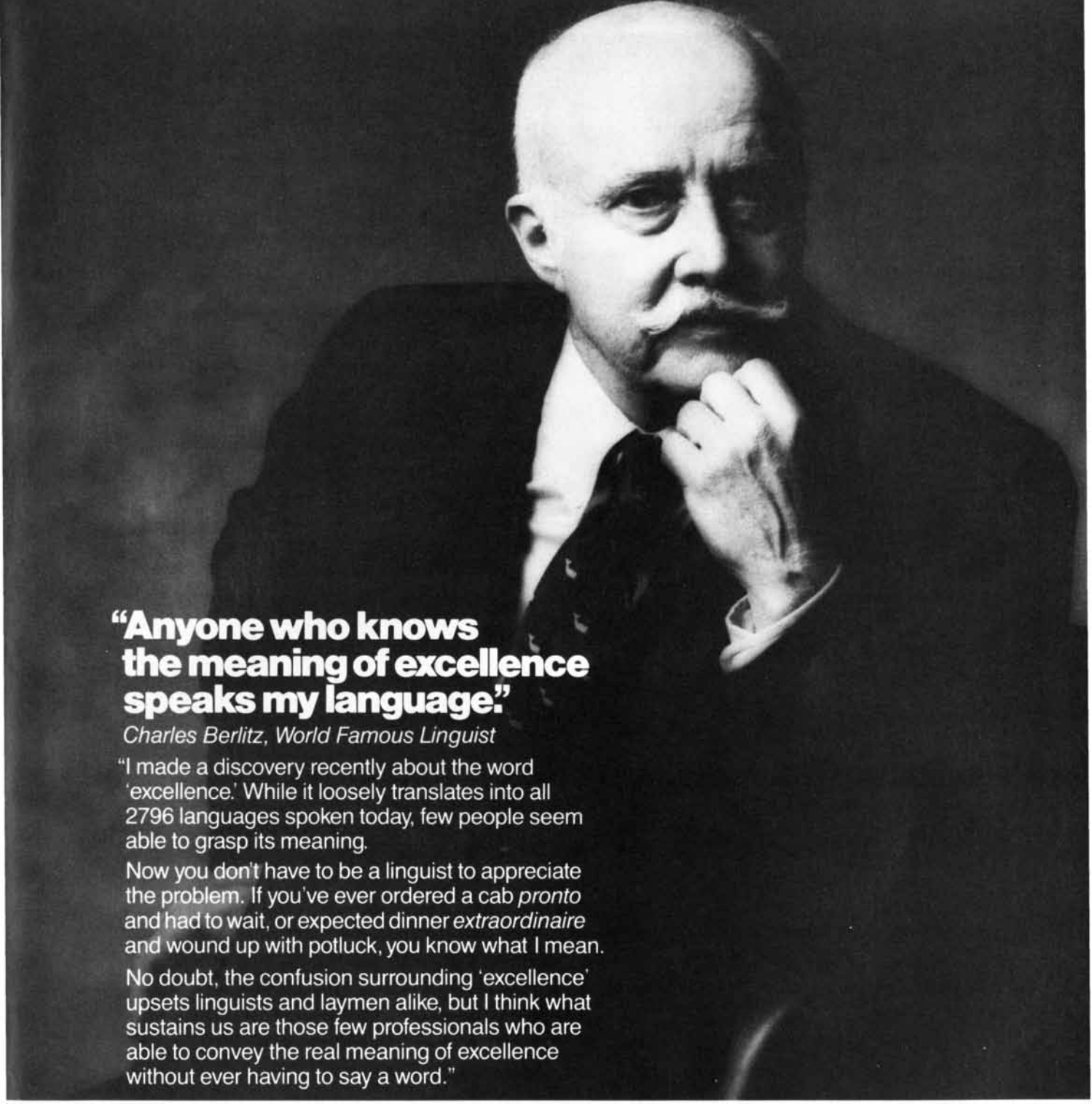
Now, during the eighth century, when Shanga was first occupied, there were no foreign settlements on the coast of East Africa. Indeed, the earliest occupation layers at Shanga appear to be completely African in style. Around a rectangular cattle corral were set a group of daub-and-mud huts; each clan grouping had its own gate into the corral, which was an area of communal activity. Although initially the residents probably were not Muslims, by sometime in the ninth century an elite group had converted to Islam, as is shown by the discovery of postholes for three small wood mosques. That the new faith was for an elite, and not for the general population, is indicated by the limited capacity of the early wood mosques, which could have held only 20 or so of the estimated population of 500.

Settlements of the same type and date are widespread on the East African coast. The northernmost is near Mogadishu in Somalia. There are several more in the Lamu archipelago (of which Pate Island is a part), two on Zanzibar and several on islands south of Zanzibar, including Mafia,



MTEPE was the traditional sailing vessel of the Swahili traders. A typical *mtepe* might have been 15 meters long. The sail was made of palm-leaf matting. The hull was built from planks of African teak, drilled and lashed with coir (rope made of the fiber of the coconut husk). The prow represents a camel's head; the flags symbolize Sultan Ali of Shungwaya, a legendary Swahili sailor, and his sons. The *mtepe* design, which may have originated in the first century A.D., was still in use in the early 20th century.

© 1987 SCIENTIFIC AMERICAN, INC



Excellence is not a foreign word at Lufthansa. It is the standard by which we work. It's the reason for our outstanding on-time arrival record, our meticulously maintained fleet and our friendly, courteous service.

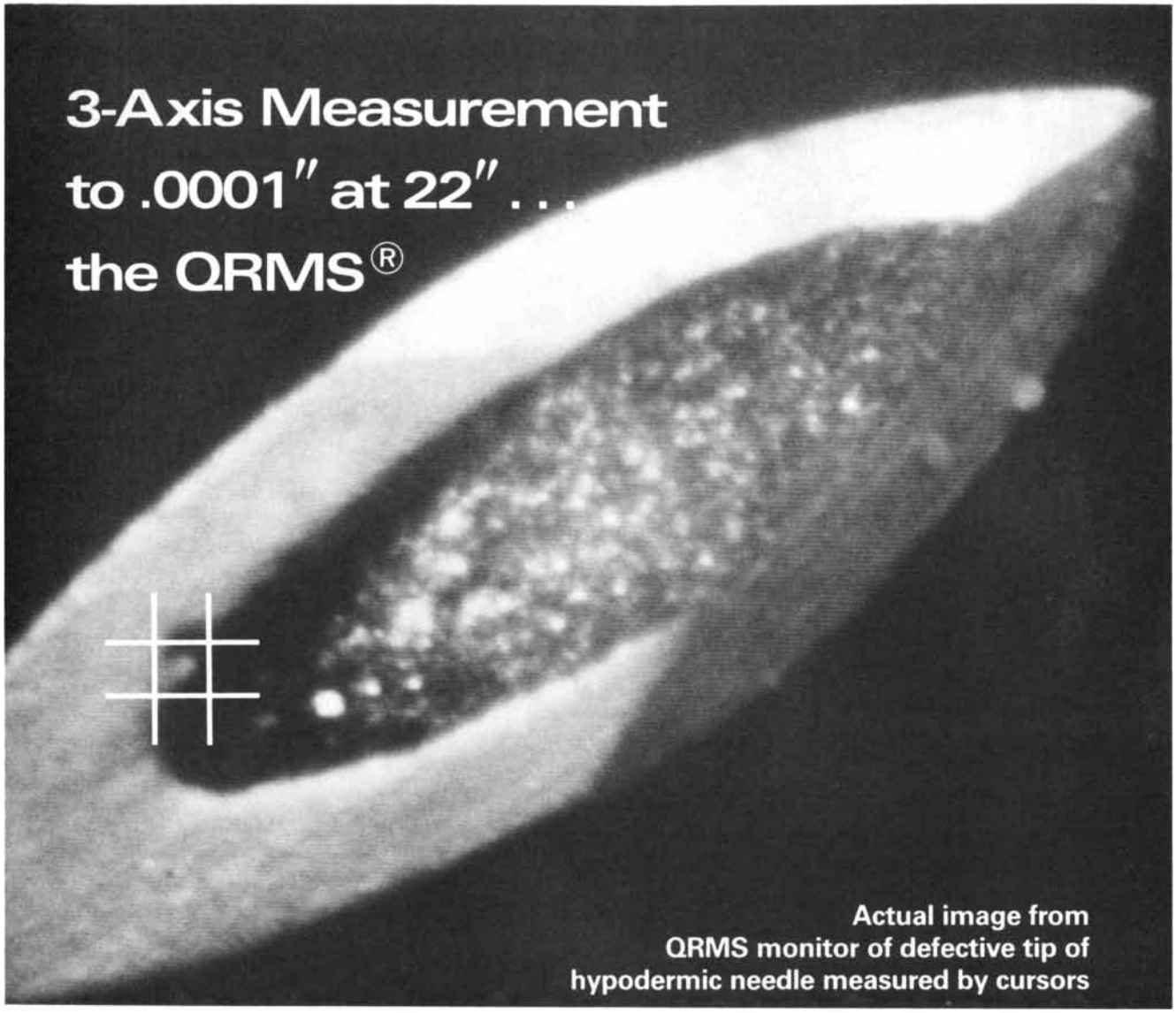
That's why in 82 countries worldwide Lufthansa means excellence.

People expect the world of us.

Lufthansa
German Airlines

Lufthansa is a participant in United's Mileage Plus program and Delta's Frequent Flyer program. See your Travel Agent for details.

© 1987 SCIENTIFIC AMERICAN, INC



Questar Corporation announces its Remote Measurement System — the QRMS — an optical AT-based, keyboard-controlled and menu-driven gauging, alignment and imaging system. Non-contact digital imaging calculates linear and angular measurement in 3 axes to .0001″ and does it with on-line monitoring in a high-speed production environment without disturbing the dynamic process. It has full image and data archival capabilities, will generate gray-scale profiles and 3-D maps, calibrates without reference, can include numeric and statistical processing of data, and is portable. The QRMS, which incorporates the award-winning Questar long-range microscope, is a powerful tool for research and production. You no longer need to stop the process, remove the part, inspect the part, replace the part. You simply

MEASURE IT . . . CONTROL IT . . . DOCUMENT IT

QUESTAR

P.O. Box 59, Dept. 217, New Hope, PA 18938 (215) 862-5277

© Questar Corporation 1987

© 1987 SCIENTIFIC AMERICAN, INC

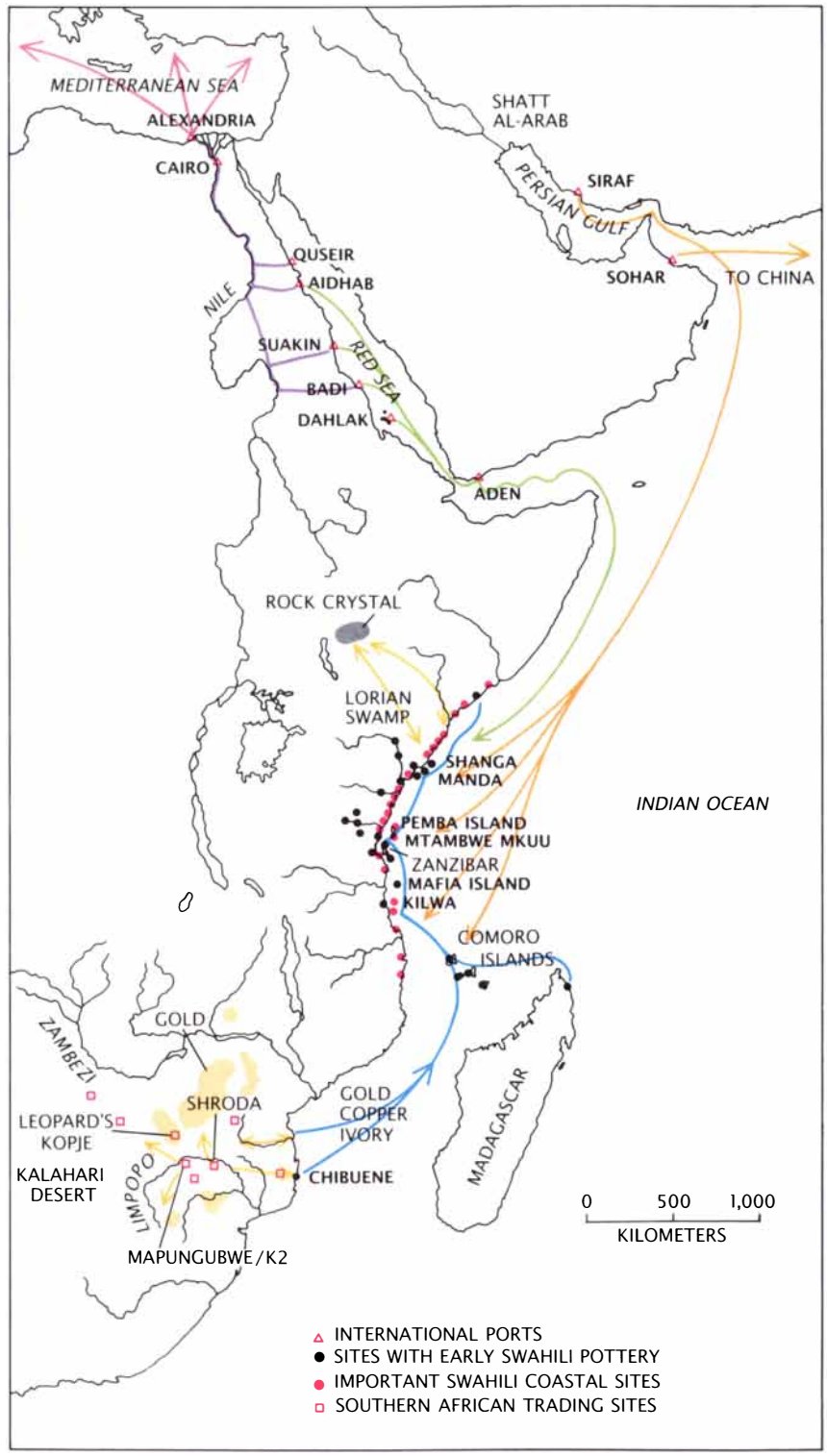
the Kilwa Islands and the Comoro Islands. The southernmost lie on the mainland of Mozambique. Such sites mark the early phase of the Swahili Corridor. Among their common features are mud houses and remains suggesting a diet of fish and wild animals. The earthenware pottery is practically identical at all the sites, suggesting a single cultural group that maintained contact by sea over extended periods.

This unified, indigenous seafaring society (the existence of which has been recognized only in the past five years) may have come into existence as early as the first century A.D. By the ninth century it had become an integral component of a large international trading network. That network, however, was not the one that led to the Mediterranean. It was a route that led to China by way of the Persian Gulf.

Among the major ports of the Persian Gulf were Siraf and Sohar, whose traders took advantage of the resources available on the East African coast to establish a trading system extending to the Middle East and China. Among African exports to China were ivory and ambergris (a waxy substance believed to come from the intestine of the sperm whale that is employed as a fixative in perfume). The bulkier products, such as wood and slaves, went to the Middle East. Wood from East African forests roofed great cities in the Middle East, perhaps even Baghdad; the slaves drained the swamps of the Shatt-al-Arab on the Euphrates. The return journey brought Chinese stoneware of the T'ang dynasty to Africa. As much as 5 percent of the pottery at Swahili sites comprises Chinese wares or glazed pottery from the Persian Gulf.

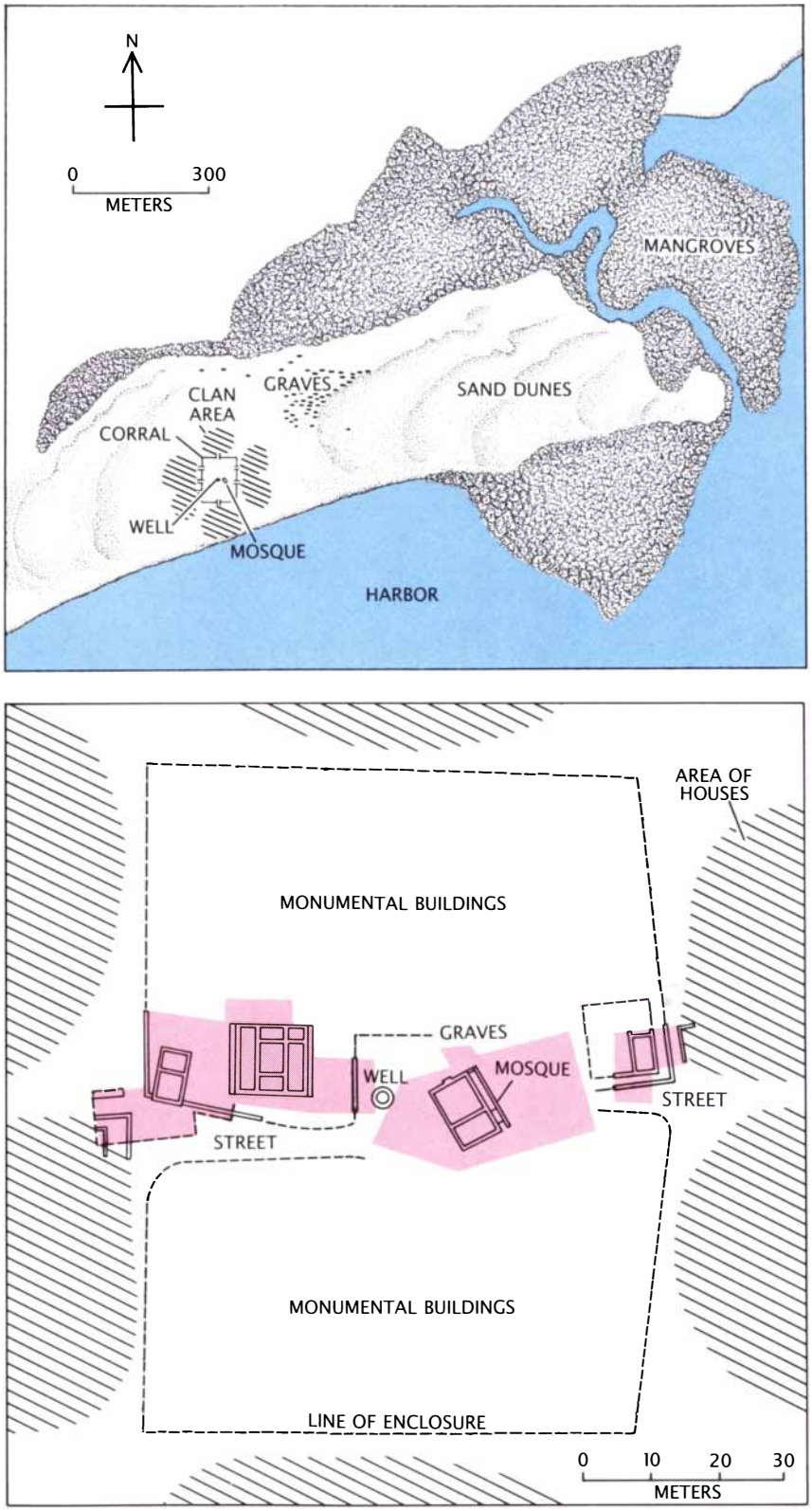
By the beginning of the 10th century the trading system centered on the Persian Gulf was in decline. The Zanj Revolt, a massive rebellion of African slaves in Mesopotamia at the end of the ninth century, reduced the scale of the African trade. Commerce between Siraf and China, which had provided the largest market for East African ivory, diminished greatly during the political instability following the ending of the T'ang dynasty in 906. Commercial decline is reflected in the archaeological record on the East African coast: many early sites failed to survive the collapse and were abandoned.

Into the vacuum created in this



SWAHILI CORRIDOR includes 3,000 kilometers of East African coastline. The corridor is dotted with Swahili coastal sites (*red circles*); at many of the sites early Swahili pottery has been found (*black circles*). In the eighth century traders from the Persian Gulf established a route linking East Africa and China (*orange*). By the 10th century they had been supplanted by other Muslim traders from the Red Sea, who set up a route connecting East Africa and the Mediterranean. Rock crystal (transparent quartz), ivory and gold were brought to the coast along inland trade routes (*yellow*). The Swahili obtained the goods at coastal ports and sailed north (*blue*), where Red Sea traders got the precious materials and carried them home (*green*). Eventually they reached the Mediterranean.

© 1987 SCIENTIFIC AMERICAN, INC



SHANGA, a typical Swahili site, is on Pate Island off the coast of Kenya. The author's excavations at Shanga revealed 25 phases of occupation lasting from the eighth through the 14th centuries. In the early phases Shanga was an Iron Age African community (*upper panel*). Mud-and-daub huts grouped in clan areas surrounded a communal corral for cattle. In about A.D. 800 the first of several wood mosques was built in the corral. In about 950 a stone complex was built on the same site (*lower panel*). The new complex included a larger stone mosque and a wall following the lines of the old corral.

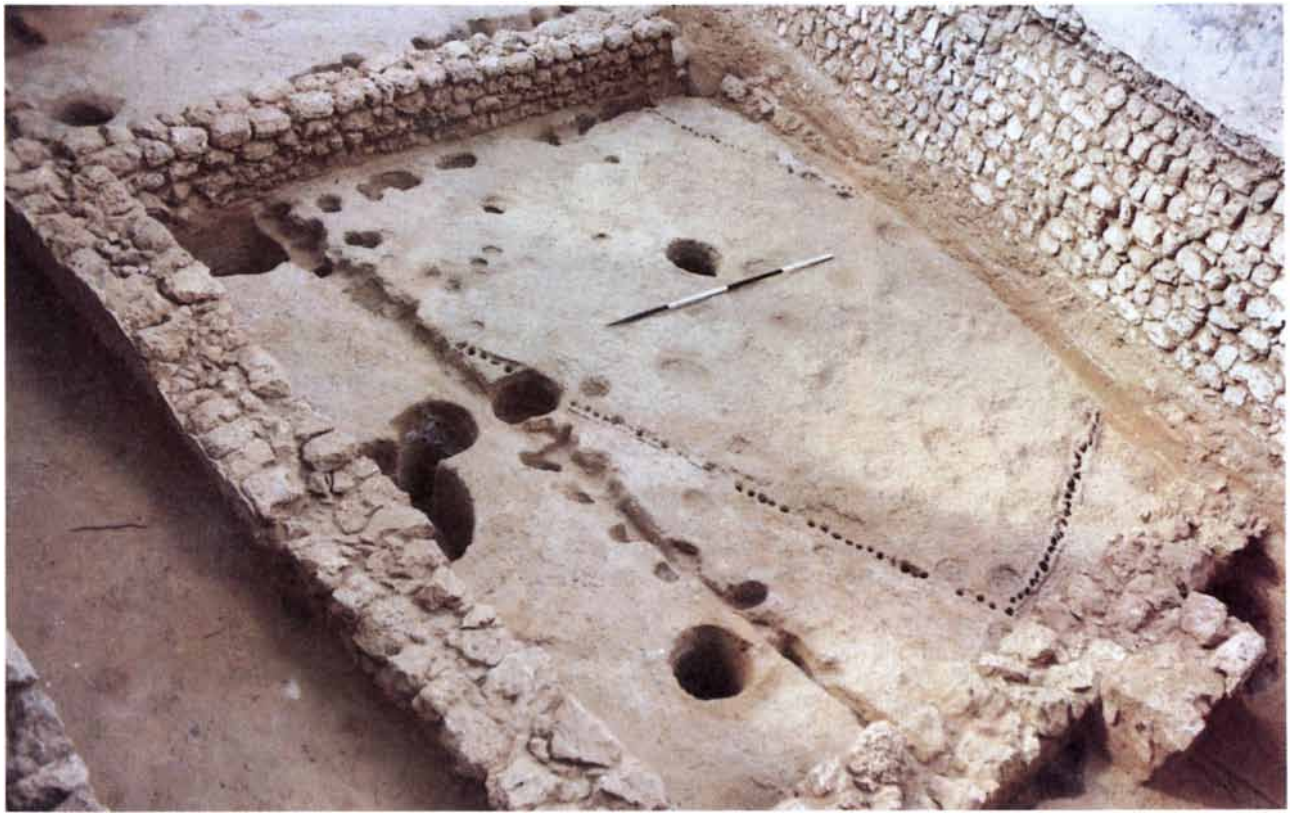
way came merchants from the Red Sea and the Gulf of Aden. Those traders, who had connections with the Mediterranean rather than with China, needed to establish permanent trading relationships with the Swahili rulers of the coast. To cement such ties they offered not only prosperity but also a package of artistic and cultural styles along with skilled craftsmen who could put such styles into practice. As a result a new way of life appeared on the East African coast. That way of life, which closely mimicked the Muslim courts of the Middle East, was adopted by the Swahili rulers who had contact with the Red Sea merchants.

The emergence of the new prosperity and its attendant way of life was first recognized in my excavation at Shanga. Its physical manifestation there was a monumental set of stone structures lying directly over the remains of the ruined daub-and-mud corral. The stone complex, erected in about 950, had walls of coral faced with fine lime plaster. The buildings were on a lavish scale and were entered by stone stairways. They were surrounded by a stone wall that closely followed the eighth-century cattle corral. In the middle of the complex stood a small stone mosque capable of holding no more than 40 people; this is the earliest stone mosque known in sub-Saharan Africa.

The stone complex bespeaks a sharp change in the way of life at Shanga in the mid-10th century. Indeed, it looks very much like a royal palace. There had been a Muslim elite at Shanga as early as the ninth century, as the finding of the three early wood mosques showed. The new buildings, however, are on a much grander scale. The enclosure wall, the central mosque and the linked stone buildings suggest a display of wealth and power by a newly prosperous ruling group. The fact that the stone enclosure follows the line of the old cattle corral indicates that the novel way of life was adopted by the indigenous people rather than imposed from outside.

Like the unified seafaring culture, the new way of life was not limited to Shanga. At the nearby site of Manda there are similar buildings, of brick as well as stone, within an enclosure that dates to about 950. Brick and stone buildings of similar date have also been found at Ungwana on the Kenyan mainland. One such house

© 1987 SCIENTIFIC AMERICAN, INC



EXCAVATIONS AT SHANGA uncovered postholes from three successive wood mosques (the earliest dating from A.D. 800) under two stone ones. The rectangle of small holes corresponds to the first wood mosque; the linear trench dotted with medium-size holes corresponds to the second; the largest postholes correspond to the third. The foundation in the foregound is from the first stone mosque, which was built in the 10th century; the wall at the rear is from the second, which was built in the 11th.

survives aboveground on Mafia, and on Kilwa several stone houses dating to the 10th or 11th century have been excavated by Neville Chittick of the British Institute in Eastern Africa. At each of these sites there seems to have been a centralized authority in contact with traders and merchants.

I have little doubt that the source of these changes was contact with the traders of the Red Sea and through them with the Mediterranean world. Some of the strongest evidence comes from the techniques that were exploited in erecting the new buildings. Their walls were made of deep-sea coral brought up by divers and worked before it hardened. Such construction is a particular feature of the coastal towns of the Red Sea. Indeed, until not long ago similar techniques were employed at ports such as Suakin on the Sudanese coast. It appears the techniques were developed during the 10th century in the Dahlak Islands and at sites such as Er Rih (ancient Badi) near Suakin. From the southern Red Sea they were brought to East Africa to enrich the courts of the rulers of the Swahili Corridor.

The clinching evidence comes not from building materials, however, but from the lifeblood of trade: money. At Mtambwe Mkuu, an 11th-century site on the island of Pemba, I found a hoard of gold and silver coins. There were more than 2,000 of them, tucked in a cloth pouch with a silver clasp that was buried in a modest timber house. According to Helen Brown of the Ashmolean Museum of Art and Archaeology in Oxford, the hoard was buried sometime after 1066, which is the date of the latest coin. Although most of the hoard is silver, 10 coins are gold, and seven of these are Fatimid dinars from Mediterranean mints. The other three are copies of Fatimid coins inscribed with bogus Arabic writing; these imitations may have been struck in East Africa. Thus the gold coins show a dramatic and direct connection with the Mediterranean.

That connection is reinforced by the silver coins, which are so small and thin that they resemble pieces of tinfoil. They were minted locally, as is shown by the names of African rulers inscribed on them. In their design, however, they clearly show the Muslim influence from the Mediterranean. The ruler's name is on the face of the coin; on the reverse is inscribed a motto rhyming with the name. The only other coins with that arrangement probably come from Fatimid Sicily in the 11th century. What is more, five genuine Fatimid coins from Sicily were found at Manda. It seems that by 950 the Swahili rulers of the Lamu archipelago possessed coins from the Mediterranean and were using them as models for minting their own.

Why did the Swahili want to emulate their Mediterranean contemporaries? One answer is that conversion to Islam may have offered the Swahili of the coast the ritual protection of a world religion, no small factor in a region where magic was a potent aspect of authority. By means of their recently acquired religion, the Swahili may have achieved a ritual equality with their merchant contacts. A shared religious (and moral) code would also have increased the security of commercial transac-

© 1987 SCIENTIFIC AMERICAN, INC

SILVER COIN gives evidence of the connection between East Africa and the Mediterranean. The coin is tiny: 11 millimeters across. Although minted in Africa, it closely follows the style of coins from mints in Sicily controlled by the Fatimids (the Muslim rulers of Spain, Sicily, North Africa and Egypt). The face (*left*) carries the name of a Swahili ruler; the reverse carries a motto that rhymes with the name. The coin is one of a hoard of 2,000 found in 1984 by the author at Mtambwe Mkuu, a Swahili site on Pemba Island.

tions, which was of advantage to both sides. Such nonmaterial benefits to the Swahili would have been reinforced both by money and by cultural innovations such as the new building techniques.

What did the Swahili have to offer the international Muslim merchants? Their most important offering was mastery of coastal trade. By the eighth century rapid sea travel was possible from Mogadishu in Somalia all the way to Mozambique, a distance of some 3,000 kilometers. The corridor was not an easy sail. The Mozambique channel, separating the mainland from Madagascar, is dangerous, and the seasonal monsoons are unreliable south of the Equator. Along this route specialized navigators were needed, and the Swahili provided them. In their teak boats with sails made of coconut matting they sailed up and down the corridor, bringing goods to the large commercial centers in the north. There the Arab traders loaded the goods into big transoceanic dhows and returned to the Red Sea.

The sailing skills of the Swahili would have counted for less had they not been able to supply the kind of goods their trading partners wanted. The problem the Swahili faced was to obtain these goods—gold, ivory and rock crystal—from the interior. To do so they exploited long-distance systems of exchange that already existed in Iron Age Africa south of the Sahara. Such exchange systems, entirely indigenous in origin, served to redistribute scarce resources such as salt. They also made it possible for groups with quite different ways of life, such as hunters, herders and farmers, to acquire one another's products. The trade involved both utilitarian goods and prestige items such as seashells.

The Swahili took the preexisting exchange network and began expanding it for their own purposes. By controlling the supply of goods needed in the interior, they developed a reliable group of inland trading partners. They obtained a monopoly of seashells, particularly cowries and conus shells, and controlled the preparation of salt from seawater. Among their number were specialists in ironworking, who supplied finished products to groups lacking those skills. Cattle and even camels moved down the coast in the teak boats, along with cloth made of African cotton or imported silk.

On the basis of such desirable items, the Swahili were able to assemble an exchange network leading to the interior. The Swahili themselves did not travel inland. Instead, as can clearly be seen in the case of the ivory-supply route, they relied on local groups. Ivory was acquired on the coast by the Swahili from trading partners who were herders; in exchange the Swahili gave finished goods, particularly cloth. The relations between the Swahili and the herders included not only trade but also military support and, from the Muslim side, the offer of magic and religious "protection."

The Swahili got ivory from herders, who in turn got it from a third link in the chain: hunting peoples. In some traditions the hunters gave one of each pair of tusks to the herders as tribute and were free to trade the second for dairy products or iron for arrowheads. Eventually all the ivory reached the coast through the complex network of trading connections. There the Swahili acquired it and took it back to their home towns in the north, where it was sold to the foreign traders at a huge profit.

Rock crystal and gold offer variations on the theme of exchange with the interior. According to recent work by Richard Wilding of the National Museums of Kenya, the most probable source of the crystal is southern Ethiopia, where the correct geologic conditions prevail for the exposure of large pieces. In the 10th century the crystal trade appears to have been in the hands of herders who may have been relatives of the modern Boran who still inhabit the region. The herders were mobile and avid for manufactured items that were not available in their own territory. They probably collected the crystal and brought it to the coast, giving it up in exchange for cloth and manufactured goods.

Once it was on the coast the crystal seems to have been worked before being transported. In 10th-century occupation levels at East African coastal sites pure rock crystal appears in the form of waste fragments as well as partially worked lumps and beads. The beads often are discards that were tossed away because they were cracked or incorrectly pierced. Such a pattern indicates that the large lumps of crystal were partially trimmed by local craftsmen on the coast and the waste was made into trading beads. The supply lasted for only about a century. By 1050 rock crystal had vanished from coastal sites in the Lamu archipelago. At about the same time the workshops in the Mediterranean ceased production in crystal and turned to substitutes such as clear glass.

Rock crystal reached the coast in the immediate hinterland of the main Swahili ports, in the north. Gold presented a more difficult problem, because the supply lay far to the south in areas reached only by a long sea journey. Gold was to be found in southern Africa, specifically between the Zambezi and Limpopo rivers, and in that region the precious metal had been mined very early. In 1969 Roger Summers of the then National Museum of Rhodesia identified more than 1,000 areas where there are ancient workings: tunnels and shafts into gold-bearing deposits. The workings

© 1987 SCIENTIFIC AMERICAN, INC

cannot be precisely dated, but some are associated with Iron Age pottery of the first millennium A.D. The few radiocarbon dates show the gold was being worked long before the arrival of the Portuguese at the end of the 15th century.

On the basis of his survey Summers estimated that a total of some 20 million ounces of gold were taken out of ancient southern Africa over about 800 years. Thus the annual yield might have amounted to more than 20,000 ounces, which would undoubtedly have had a major impact on the world economy. In the light of such figures it is no surprise that the Swahili placed great importance on the gold trade.

The coastline of southern Africa constituted the terminus of the Swahili Corridor. There the merchants coming from the north obtained not only gold but also ivory, as a number of recent finds suggest. A typical Swahili port on the southern coast is Chibuene, which was excavated by Paul Sinclair for Eduardo Mondlane University in Mozambique. Chibuene was not a permanent settlement but a seasonal encampment where Swahili traders met traders from far in the interior. The cultural unity of the corridor is suggested by the fact that although Chibuene is 2,500 kilometers from Shanga, the local earthenware pottery and the imported glazed jars and bowls are identical with those found in the Lamu archipelago.

From ports such as Chibuene trade routes penetrated the interior, reaching the goldfields and the elephant herds. Recent discoveries reveal that the inland routes extended 1,000 kilometers from the coast into the Kalahari Desert. There the San hunters exchanged ivory for glass beads. That trade had begun by the seventh century, but the richest sites are later ones. At Shroda and also at a site designated K2, which has been dated to about 960 by radiocarbon methods, large quantities of glass beads and refuse from ivory working have been discovered. At nearby Mapungubwe rich royal burials were found, including one with a spectacular gold rhinoceros. The richness of those sites reflects their location on the southern fringe of the goldfields in a good spot for controlling trade with the coast.

The routes by which gold, rock crystal and ivory reached the coast complete our picture of the trade route extending from southern Africa to Sicily and beyond. The entire network developed rapidly in the second half of the 10th century and lasted well into the 16th. It is not likely that such a complex system was set up simply in response to the new demand from Europe. Patterns of local demand and social organization were equally important. The Swahili were already trading on the coast, gold was being mined in southern Africa and indigenous exchange systems were in place. It was the linking of such components that established the Mediterranean route.

The main explanatory problem for an archaeologist or a historian is to show whether the connections were achieved fortuitously or by the initiative of one group in the network. It does seem improbable that such a long, dangerous route was established by chance. The evidence suggests that only one group was in a position to appreciate the large profits that could be made and also had contacts extending from the Mediterranean to southern Africa: the Swahili. It is salutary to think the first flowering of medieval culture in Europe rested on the skills of the sailors of the Swahili Corridor.

GOLD COINS FROM MTAMBWE MKUU include Fatimid dinars and imitations that are probably African. The dinar was the main unit of currency in the Islamic world. The coins in the upper rows are from Fatimid mints in Tunisia, Egypt, Damascus and Tyre. Those in the bottom row are imitations, perhaps made by Swahili rulers to emulate their Mediterranean counterparts.

© 1987 SCIENTIFIC AMERICAN, INC

How Children Learn Words

The key is to see words in intelligible contexts. A dictionary is often misunderstood, but an interactive video display can mobilize the natural ability of a child to learn from context

by George A. Miller and Patricia M. Gildea

Listening to a child who is just learning to talk, one is most aware of the child's limited command of the language. What one tends to overlook is the sheer magnitude of the child's achievement. Simply learning the vocabulary is an enormous undertaking. The fact is that for many years after starting to talk a child learns new words at a rate of more than 10 per day! Yet little is known about how children do it. Certainly they do not do it by memorizing dictionary entries. Our findings and those of other workers suggest that formal efforts to build vocabulary by sending children to the dictionary are less effective than most parents and teachers believe. We are exploring the possibility that a computer program providing lexical information about new words encountered in the context of a story might be more effective.

When adults set out to learn a new language, they know what is in store. They realize they will have to learn a new pronunciation, a new grammar, a new vocabulary and a new style of using language. They know they will have to spend many hours every day for years before they can call themselves fluent in the new language. They also know, however, that they will be able to rely on teachers to explain, in their first language, everything they need to learn about the second language.

How different it is for infants. Having no language, they cannot be told what they need to learn. Yet by the age of three they will have mastered the basic structure of their native language and will be well on their way to communicative competence. Acquiring their first language is the most impressive intellectual feat many people will ever perform.

Students of how children learn language generally agree that the most remarkable aspect of this feat is the rapid acquisition of grammar. Nevertheless, the ability of children to conform to grammatical rules is only slightly more wonderful than their ability to learn new words.

How many words must one know in order to use English effectively? The answer depends on several variables, including the definition of "word." For the purpose of counting, a word can be defined as the kind of lexical unit a person has to learn; all the derivative and compound forms that are merely morphological variations on the conceptual theme would not be counted as separate words. For example, *write* is a word and its morphological variants (*writes, writ, wrote, written, writing, writer* and so on) are relatives in the same family. If such a family is counted as a single word and knowing a word is defined as being able to recognize which of four definitions is closest to the meaning, the reading vocabulary of the average high school graduate should consist of about 40,000 words. If all the proper names of people and places and all the idiomatic expressions are also counted as words, that estimate would have to be doubled.

This figure says something about the ability of children to learn words. If the average high school graduate is 17 years old, the 80,000 words must have been learned over a period of 16 years. Hence the average child learns at the rate of 5,000 words per year, or about 13 per day. Children with large vocabularies probably pick up new words at twice that rate. Clearly a learning process of great complexity goes on at a rapid rate in every normal child.

No one teaches children 13 or more words a day. Children must have a special talent for this kind of learning. Some valuable hints as to how they do it were uncovered a decade ago by Susan Carey and Elsa J. Bartlett, who were then at Rockefeller University. They worked with the names of colors. First they established that a group of three-year-olds did not know the color olive. Most of the children called it green and some of them called it brown.

Carey and Bartlett taught the children a nonsense name for olive—a name they would not have heard anywhere else. They took two cafeteria trays and painted one tray olive and the other blue. Each child was then told casually, "Hand me the chromium tray. Not the blue one, the chromium one." The child would pause and perhaps point to the olive tray. "This one?" "Yes, that one. Thank you."

A week later, with no further guidance, the children were again asked to name the colors. When olive was presented, they paused. They did not remember *chromium,* but now they knew that this color was not called green or brown. A single exposure was enough to begin a reorganization of their color lexicon.

This simple experiment demonstrated some important points about how children learn words. First, in order to learn a word a child must be able to associate its sound with its meaning. Mastering the mechanics of uttering and recognizing a word and mastering the concept that it expresses are separate learning processes. After their experience with the trays the children knew that olive has a special name—that it is not called green or brown—but they did not remember the particular spoken sound associated with that perceived color. Many repetitions may be necessary before the sound of a new word becomes familiar.

© 1987 SCIENTIFIC AMERICAN, INC

Second, a child's appreciation of the meaning of a word seems to grow in two stages, one rapid and the other much slower. Children are quick to notice new words and to assign them to broad semantic categories. After hearing *chromium* just once the three-year-olds assigned it to the semantic field of color names. Children are able to keep such fields separate even before they know what the individual words mean. Asked the color of something, they may respond with almost any color term at random, but they never answer *round* or *five* or *lunch*.

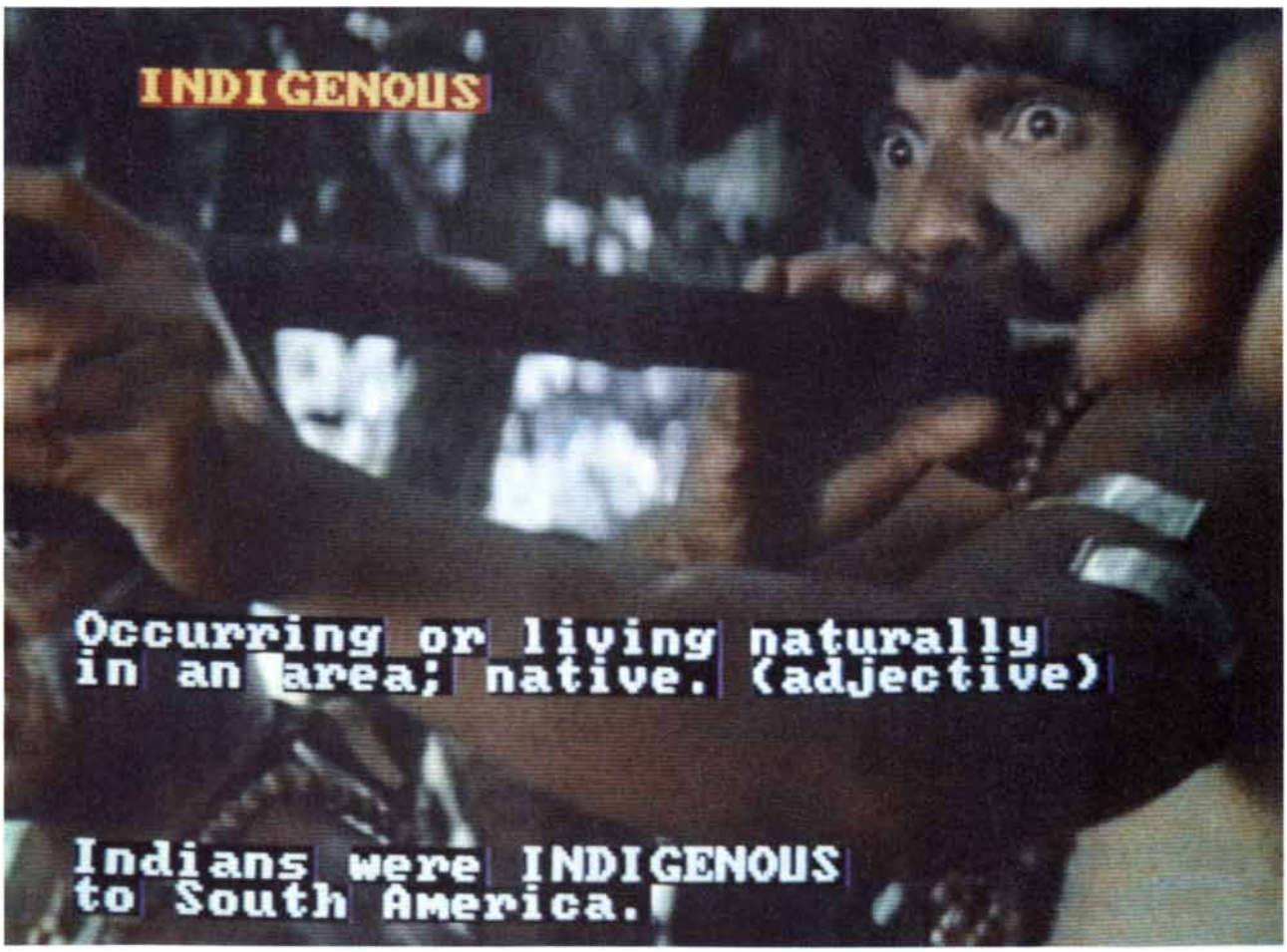
The slow stage entails working out the distinctions among words within a semantic category. A child who has correctly assigned *red, green, yellow* and *blue* to the semantic field of color terms still has to learn the differences between and relations among those words. This stage ordinarily takes much longer than the first and may never be completely finished; some adults, for example, correctly assign *delphinium* and *calceolaria* to the semantic field of flowering-plant names but have not learned what plants the words denote and cannot identify the flowers on sight. At any given time many words will be in this intermediate state in which they are known and categorized but still not distinguished from one another.

A related aspect of word learning by preschoolers that has attracted wide attention is called overextension. For example, a small child learning the word *apple* may apply it to a tomato. *Apple* is thought to mean, say, round, red and of a certain size; without further qualification those attributes define ripe tomatoes as well as ripe apples. Overextension can occur when the child's conception of a word's meaning is incomplete.

The opposite error also occurs, but it is revealed only by special questioning. For example, a child who thinks that being round, red and of a certain size defines *apple* might fail to use *apple* to refer to green or yellow apples. The only way to identify such an underextension is to show the child green or yellow apples and ask what they are called.

The ability of preschoolers to soak up words has attracted increasing attention in recent years. Much more is known about it than was known when Carey and Bartlett did their pioneering study with color names. The word-learning process becomes even more complex, however, during the school years.

In the early grades schoolchildren are expected to learn to read and write. At first they read and write familiar words they have already learned by means of conversation. In about the fourth grade they begin to see written words they have not



COMPUTERIZED TUTORING is being tested by the authors as an improvement over dictionaries in helping children to grasp the meaning of an unfamiliar word. The children read a text describing an episode from a motion picture they have just seen, in this case *Raiders of the Lost Ark*. The text contains specially marked words, such as *indigenous*, that the children are expected to learn. Interacting with a video display, the children can ask for information about the word in any or all of three forms: definitions, sentences and pictures. The aim is to present information about a word when the child is motivated to learn the word.

© 1987 SCIENTIFIC AMERICAN, INC

heard in conversation. At this point it is generally assumed that something special must be done to teach children these unfamiliar words.

This educational assumption runs into serious problems. Although children can recognize that they have not seen a word before, learning it well enough to use it correctly and to recognize it automatically is a slow process. Indeed, learning a new word entails so much conceptual clarification and phonological drill that there simply is not enough classroom time to teach more than 100 or 200 words a year in this way. Since learning runs so far ahead of teaching—some 5,000 words learned in a year compared with 200 taught—it is hard to avoid the question: How do schoolchildren learn so much more than they are taught?

Many words are acquired through reading. Children learn words at school in the same way as they do at home: by observing how the words are used in intelligible contexts. The difference is that the academic environment depends more on written contexts. Both public opinion and scientific evidence are converging on the view that the best way to facilitate vocabulary growth in schoolchildren is to have them read as much as possible.

Learning words by reading them in context is effective but not efficient. Some contexts are uninformative, others misleading. If the word in question expresses an unfamiliar concept, a single context of use will seldom support more than one hypothesis about the word's meaning. In order for reading to have any substantial effect on vocabulary a great deal of reading must be done.

How much? A child who spent 50 minutes of every school day reading at, say, 200 words per minute would read one million words in a 100-day school year. A million running words of English prose would typically contain no more than 50,000 distinct word types, representing roughly 10,000 word families. Schoolbooks would probably contain fewer different words. Even among 10,000 different words, it is unlikely that more than 1,000 would be totally new lexical items. Since multiple encounters are required in order to learn a new word, it is clear that reading one million words per year is not enough. In order to account for a growth rate of 5,000 words in a year it seems necessary to think about continued learning from conversational interactions supplemented by reading several million words per year. Indeed, children who read little outside the classroom generally do poorly on vocabulary tests.

The fact that children learn many more words than anyone has time to teach them also carries implications for the role of teachers in this learning process. Learning new words from purely literary contexts of use—from the contexts provided on the printed page—is harder than learn-

TRAY EXPERIMENT showed how quickly preschool children assign new words to semantic categories. A decade ago Susan Carey and Elsa J. Bartlett, who were then at Rockefeller University, established that a group of three-year-old children did not know the name for the color olive; they called it green or brown. The experimenters painted one tray blue and another one olive and asked each child to "hand me the *chromium* tray, not the blue one." A week later the children were asked to name the colors. They did not remember *chromium* but now knew the color was not called green or brown. A single exposure was enough to cause them to reorganize their semantic field of color terms. The photograph is of a reenactment of the original experiment.

© 1987 SCIENTIFIC AMERICAN, INC

ing them through interaction with a person. In conversation it is usually possible to ask the speaker what an unfamiliar word means. Moreover, in most conversations visual information supplements the linguistic information. Such help is missing from the printed page.

Given this additional difficulty, it seems reasonable to ask teachers to help children to be more efficient in learning new words from context. If they cannot teach all the words children need to know, perhaps teachers could help their students learn how to work out such things for themselves.

One way to figure out what an unfamiliar word means is to use a dictionary. In about the fourth grade, therefore, most schools begin to teach dictionary skills: spelling, alphabetizing, pronunciation, parts of speech and a little morphology and etymology. The idea, which is perfectly reasonable, is that children should learn how to find unfamiliar words in a dictionary and how to understand what they read there.

One trouble with this approach is that most healthy, right-minded children have a strong aversion to dictionaries. There may be good reason. We have looked at some of the tasks teachers assign in order to get students to use dictionaries. In our opinion these exercises do not merit the faith that teachers and parents have put in them.

Two tasks are often assigned when children are being taught how to use a dictionary. One task entails disambiguation: the child is given a sentence that contains an ambiguous word—a word with two or more senses—and told to find it in the dictionary and to decide which sense the author of the sentence had in mind. The other task calls for production: the child is given a word and told to look it up in the dictionary and to write a sentence incorporating it. On the face of it both tasks look as though they should be instructive. It is therefore surprising to discover how ineffectual they are.

Learning from a dictionary requires considerable sophistication. Interrupting your reading to find an unfamiliar word in an alphabetical list, all the while keeping the original context in mind so that you can compare it with the alternative senses given in the dictionary, then selecting the sense that is most appropriate in the original context—that is a high-level cognitive task. It should not be surprising that children are not good at it. Even when most of the complications are removed, children are still not good at it. On a simplified disambiguation task, in which fourth-grade students were given just two senses and asked to choose the one that was intended in a particular sentence, the students did little better than chance.

The second task, producing a sentence incorporating a new word, has the virtue of requiring the student to use the word and so, presumably, to think about its meaning. We have studied this production task extensively. After reading several thousand sentences that were written by children in the fifth and sixth grades we have concluded that it too is a waste of time.

Typical of the curious sentences we encountered was "Mrs. Morrow stimulated the soup." It illustrates the most frequent kind of error made by children in that age range. If they already know the word, their sentences are usually all right. If the word is unfamiliar, however, the results are often mystifying. In order to understand what the child did, you have to read carefully the same dictionary definitions the child read. The child who looked up *stimulate* found *stir up* among the definitions.

The example provides a key to what happens when children consult a dictionary. They find the unfamiliar word and then look for a familiar word or phrase among the definitions. Next they compose a sentence using the familiar word or phrase and substitute the new word for it. One of our favorite examples came from a fifth-grader who looked up the unfamiliar word *erode,* found the familiar

OVEREXTENSION in the use of words appears among preschool children when their understanding of a word is incomplete. A child whose understanding of *apple* does not extend beyond the fact that the object is round, red and of a certain size may call a tomato an apple, because without qualification those attributes also define a ripe tomato.

UNDEREXTENSION also appears, but it is revealed only by questioning. A child who thinks being round, red and of a certain size defines *apple* may not apply the word to green or yellow apples. One can find out only by asking what such apples are called.

© 1987 SCIENTIFIC AMERICAN, INC

phrases *eat out* and *eat away* in the definition and thought of the sentence "Our family eats out a lot." She then substituted *erode* for *eats out;* the resulting sentence was "Our family erodes a lot."

If children are so good at learning new words when they hear or see them used in context, why do they have trouble learning new words when they see them in a dictionary? We decided to look more closely at what goes on when an unfamiliar word is encountered in the context of a typical sentence. A preliminary study indicated that children can write better sentences when they are given a model sentence employing the word than when they are given a definition of the word. Since many of the sentences they wrote were patterned on the models, this result could not be interpreted to mean that the children learned more about the meaning of a word from illustrative sentences than they learned from definitions. Nevertheless, the observation was encouraging, and we pressed on.

The next step was simple: if one example is good, three should be better. When we made this comparison, however, we found that the number of examples made little difference. The acceptability ratings of sentences written after seeing one model sentence were the same as the ratings of sentences written on the basis of three examples.

That observation made us think again about what was going on. Apparently three unrelated sentences are hard for children to integrate, and so they simply focus on one of three examples and ignore the others. This behavior resembles what children do in reading dictionary definitions.

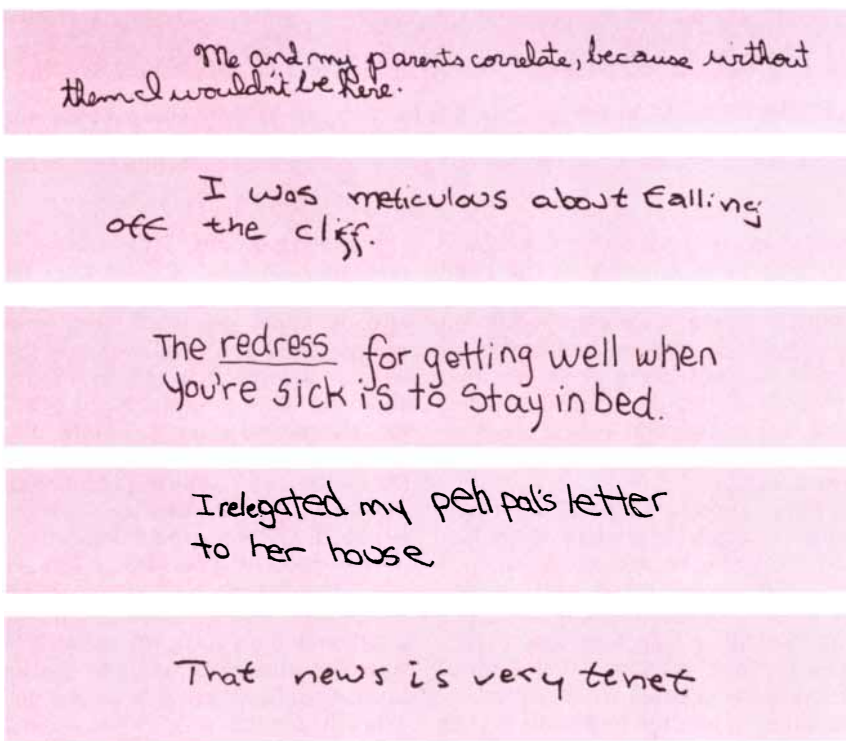
We were surprised by one result, although perhaps in retrospect we should have expected it. Mistakes resembling simple substitutions appeared even when model sentences were given instead of dictionary definitions. For example, given the model sentence "The king's brother tried to usurp the throne" to define the unfamiliar word *usurp,* the children wrote such sentences as "The blue chair was usurped from the room," "Don't try to usurp that tape from the store," "The thief tried to usurp the money from the safe" and so on. They had gathered from the model sentence that *usurp* means *take,* and so they composed sentences using *take* and then substituted *usurp* for it.

Children can appreciate at least part of the meaning of an unfamiliar word from its context, as in the case of *take* as one component of the meaning of *usurp.* Just as younger children may overextend *apple* because they know only part of its meaning, so this partial definition of *usurp* resulted in its being overextended. That is to say, if *usurp* is incompletely defined as *take,* it can be said of anything takable: chairs, tape, money or whatever. When it is seen from this perspective, the behavior of these children in the fifth and sixth grades is merely a later stage in the development of a word-learning process employed by preschool children.

The substitution strategy therefore seems to be quite general. In the context of a model sentence, however, something more than a simple substitution error appears. The children cannot search through an illustrative sentence for a familiar word as they could in a dictionary definition. First they must abstract a familiar concept from the context of the unfamiliar word. Only then can they apply the substitution rule.

Might there be a better way to foster the growth of vocabulary? What we and others have found out about the word-learning process will support some plausible suggestions. Put at the front of your mind the idea that a teacher's best friend in this endeavor is the student's motivation to discover meaning in linguistic messages. Then the problems with the traditional modes of instruction will begin to make sense. Drill on arbitrarily preselected lists of words seldom takes place at a time when the student feels a need to know those words; it fails to draw on the natural motivation for learning the associations between word and meaning. Learning through reading faces the opposite problem: not enough information about the word is available at the moment the student is motivated to learn its meaning.

What is needed is reading, which can make students curious about unfamiliar words, supplemented by immediate information about the meaning and use of those words. The important thing is to provide the information while the reader still wants it. Dictionaries are too slow. Recourse to a dictionary may help a mature and well-motivated student, but for the average child in the elementary grades it is likely to compound interruption with misunderstood information. A human tutor—someone immediately available to detect and resolve lexical misunder-



MYSTIFYING SENTENCES are often written by schoolchildren when their grasp of an unfamiliar word is incomplete. Here are examples in the handwriting of children in the fifth and sixth grades. The illustration on the opposite page reveals what was going on.

© 1987 SCIENTIFIC AMERICAN, INC

standings—would be much better than a dictionary.

Given the shortage of attentive tutors to sit at every young reader's elbow, it is natural to wonder how much of the tutoring task might be carried out by a suitably programmed computer. For example, suppose reading material was presented to the student by a computer that had been programmed to answer questions about the meanings of all the words in the material. No alphabetical search would be needed: the student would simply point to a word and information about it would appear. No sophisticated disambiguation would be necessary: the computer would know in advance which particular sense of a word was appropriate in the context. Indeed, no definition would be necessary: the phrase or sentence containing the word could be rephrased to show what the word meant in the context.

As a case in point, imagine what such a computer might do with *erode* and *usurp.* It might present a text containing the sentence "The president's popularity was eroded by his bad relations with Congress." If the student asked for information about *erode,* the computer might state: "Things can erode; when soil is eroded by rain or wind, it breaks up and so is slowly destroyed and removed. Someone's power or authority can erode too, being slowly destroyed or removed by unfavorable developments. That kind of erosion is meant in the sentence about the president."

Suppose that for *usurp* the computer presented a text containing the sentence "The king's brother failed in his effort to usurp the throne." Asked for information, the computer might say: "When you usurp a title, job or position from someone else, you seize it or take it away even though you have no right to it. In the sentence about the king's brother, *throne* means not just the piece of furniture the king sits on; it also stands as a symbol of the king's authority."

Providing such explanations almost instantly is well within the range of currently available computer technology. It is even possible to add a voice that pronounces the target word and explains it, or to show pictures indicating what the word denotes in the context.

We are exploring some of these possibilities with a setup in which children in the fifth and sixth grades interact with a video display. They are asked to read a text that describes an episode from a motion picture they have just seen. Included in the text are certain marked words the reader is expected to learn. When one of them comes up, the child can ask for information about its meaning in any or all of three forms: definitions, sentences and pictures.

For some children illustrative sentences are more informative than definitions or pictures. When such children are given a definition, they read it and quickly return to the story. When they are given a sentence that is relevant to the story and uses the word in the same context, they interpret it as a puzzle to be solved. They spend more time thinking about the meaning of the word and remember it better a week later.

We found that providing information when it is wanted can significantly improve the children's grasp of unfamiliar words, as is demonstrated by their ability to recognize the meanings and to write acceptable sentences incorporating the words. The results reinforce our belief that much can be done with computers to make learning words easier.

DEFINITION	EXCERPT	RESPONSE
correlate 1. be related one to the other: *The diameter and circumference of a circle correlate.* 2. put into relation: *Try to correlate your knowledge of history with your knowledge of geography.* v., correlated, correlating.	be related	Me and my parents correlate, because without them I wouldn't be here.
meticulous very careful or too particular about small details. *adj.*	very careful	I was meticulous about falling off the cliff.
redress 1. set right; repair; remedy: *King Arthur tried to redress wrongs in his kingdom.* 2. reparation; setting right: *Any man deserves redress if he has been injured unfairly. v., n.*	remedy	The redress for getting well when you're sick is to stay in bed.
relegate 1. send away, usually to a lower position or condition: *to relegate a dress to the rag bag.* 2. send into exile; banish. 3. hand over (a matter, task, etc.). *v.*	send away	I relegated my pen pal's letter to her house.
tenet opinion, belief, principle, or doctrine held as true. *n.*	true	That news is very tenet.

DICTIONARY DEFINITIONS read by the children who wrote the sentences in the illustration on the opposite page appear at the left here. When the word is unfamiliar, the child usually abstracts some familiar concept (*middle*) from the definition, composes a sentence embodying that concept and then substitutes the new word, such as *correlate.*

© 1987 SCIENTIFIC AMERICAN, INC

Coal-fired Power Plants for the Future

Existing plants can be improved to generate electricity more cleanly and economically. Future plants incorporating technologies expressly developed to minimize both costs and emissions will do even better

by Richard E. Balzhiser and Kurt E. Yeager

The invention of the incandescent light bulb by Thomas A. Edison in 1879 created a demand for a cheap, readily available fuel with which to generate large amounts of electric power. Coal seemed to fit the bill, and it fueled the earliest power stations (which were set up at the end of the 19th century by Edison himself). As more power plants were constructed throughout the country, the reliance on coal increased. Since World War I coal-fired power plants have accounted for about half of the electricity produced in the U.S. each year. In 1986 such plants had a combined generating capacity of 289,000 megawatts and consumed 83 percent of the nearly 900 million tons of coal mined in the country that year. Given the uncertainty in the future growth of nuclear power and in the supply of oil and natural gas, coal-fired power plants could well provide up to 70 percent of the electric power in the U.S. by the end of the century.

Yet in spite of the fact that coal has long been a source of electricity and may remain one for many years (coal represents about 80 percent of U.S. fossil-fuel reserves), it has actually never been the most desirable fossil fuel for power plants. Coal contains less energy per unit of weight than natural gas or oil; it is difficult to transport, and it is associated with a host of environmental issues, among them acid rain. Since the late 1960's problems of emission control and waste disposal have sharply reduced the appeal of coal-fired power plants. The cost of ameliorating these environmental problems, along with the rising cost of building a facility as large and complex as a coal-fired power plant, have also made such plants less attractive from a purely economic perspective.

Changes in the technological base of coal-fired power plants could restore their attractiveness, however. Whereas some of these changes are evolutionary and are intended mainly to increase the productivity of existing plants, completely new technologies for burning coal cleanly are also being developed. These technologies seek to make cost-effective environmental control, flexibility in the type of coal burned and reduced construction time inherent features of future coal-fired power plants.

To appreciate the significance of such advances in coal-combustion technology, a brief review of the operation of a conventional coal-fired power plant is in order. A modern power plant burns coal in a boiler: a huge box-shaped vessel whose inner walls contain tubes in which water is converted into steam. The coal is finely ground before it is injected into the boiler so that it burns with the high efficiency of a combustible gas. A typical large boiler consumes 500 tons of pulverized coal every hour in order to produce 6.4 million pounds of steam—enough to generate a million kilowatt-hours of electricity. At the same time the boiler gives off about 3.5 million cubic feet of flue gas, or exhaust.

The steam passes through a superheater, where its temperature and pressure are increased, before it drives a high-pressure turbine. The rotation of the turbine shaft provides the mechanical energy that is converted into electricity by a generator. To increase the overall energy-conversion efficiency, the steam leaving the turbine is usually reheated and sent back to drive one or two lower-pressure steam turbines before it is cooled, condensed and pumped (as water) to the boiler to begin the cycle once again.

In addition to coal-feed mechanisms, boilers, turbines, condensers, pumps and generators, power plants also depend on elaborate cooling, emission-control and waste-handling systems. All these main and adjunct systems must be designed to operate at high reliability for a lifetime of 40 or more years under load conditions that may vary daily between 20 percent of the plant's design capacity and its maximum power output. The capital costs of the equipment in a typical 1,000-megawatt plant can easily exceed $1 billion.

The efficiency with which the heat released from the burning of coal is converted into electricity, which was as low as 5 percent before 1900, had reached about 40 percent by 1967. In other words, in the space of about 70 years an eightfold reduction in coal consumption per unit of generated electricity was achieved. This progress was reflected in the cost of a kilowatt of new generating capacity: in 1920 it was about $350 (in 1967 dollars) and in 1967 it had dropped to $130. The cost of residential electricity service also dropped from more than 25 cents per kilowatt-hour to two cents per kilowatt-hour over the same period.

Beginning in the 1960's, however, the pace of improvement began to slow. This trend suggested that conventional power-plant designs had approached the inherent limits set by the laws of thermodynamics and the properties of the materials from which boilers and steam turbines were made. By the 1970's the dimin-

© 1987 SCIENTIFIC AMERICAN, INC

ishing technical returns were also joined by new economic and institutional forces, namely rapidly escalating capital costs, slow growth of demand, stricter emission-control requirements and extended licensing and construction schedules. The result has been an abrupt end to the historic trend of declining real cost for electricity from coal. Indeed, a kilowatt of new generating capacity costs more today (in constant dollars) than it did in 1920.

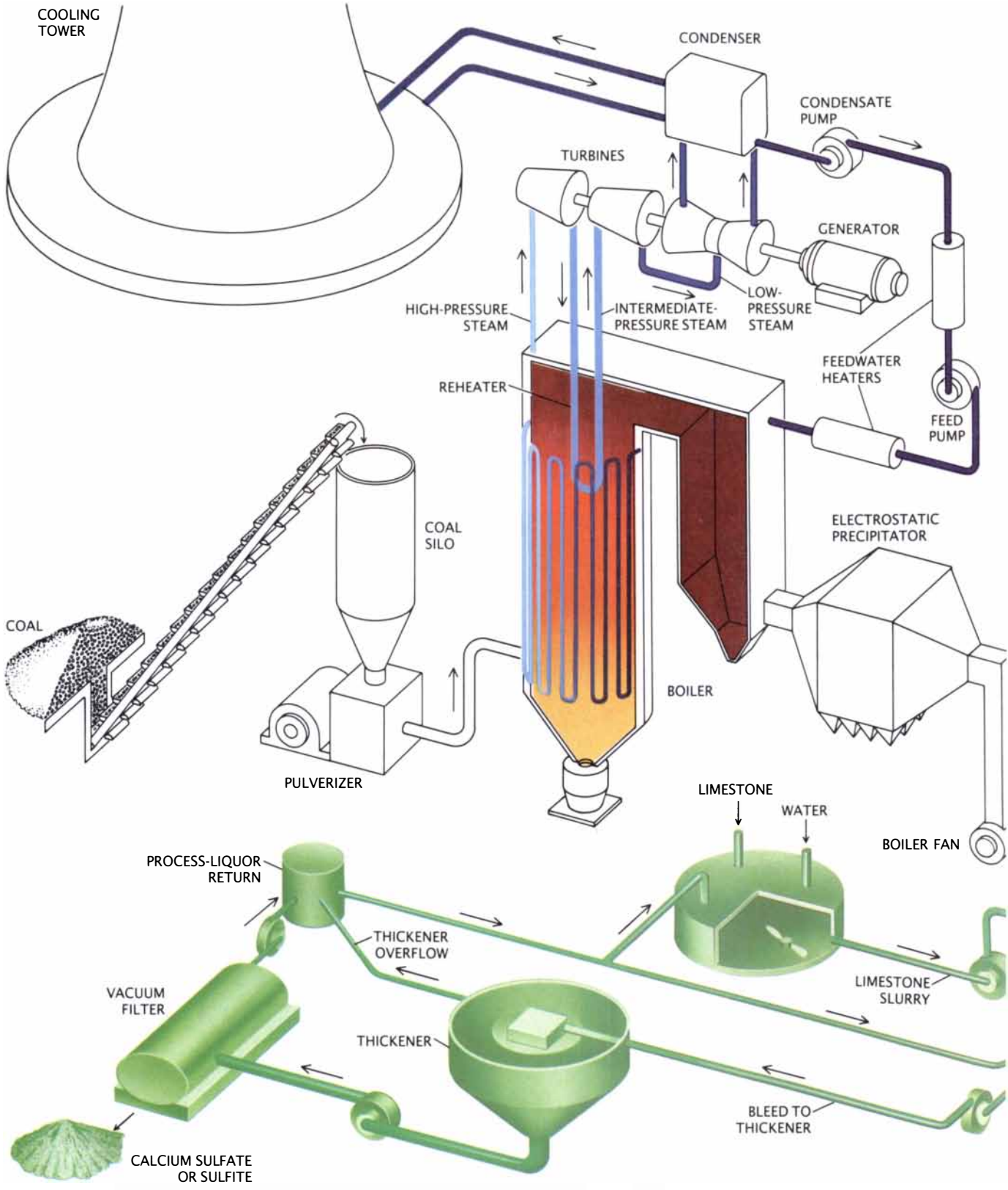
Measures to control gaseous, liquid and solid wastes have had the greatest impact on the cost of coal-fired power plants in the past 20 years. A new plant's pollution-control systems now account for as much as 40 percent of the capital outlay and 35 percent of operational costs. The most technically and economically visible element of emission control in today's pulverized-coal plants is the flue-gas-desulfurization system, often called a wet scrubber. The scrubber removes the oxides of sulfur that are a major gaseous pollutant resulting from the burning of coal.

Wet scrubbing is a simple concept but is difficult and expensive to carry out in practice. An alkaline substance, usually lime or limestone, is mixed with water to form a slurry that is sprayed into the flue gas. The sulfur oxides in the flue gas are absorbed by the slurry and precipitate out of the liquid as inert calcium sulfite or calcium sulfate (gypsum). Gypsum, if it is pure enough, can be easily disposed of or sold as a building material. Alternative but more expensive scrubbing systems are able to transform the precipitates into sulfuric acid or elemental sulfur, which can then be sold on the chemical market at a higher price. Scrubbers have been required on all pulverized-coal power plants whose construction began after 1978. As a result the U.S. electric-utility industry is operating more wet-scrubber systems today than the rest of the world combined.

In a new plant the scrubber typically costs between $150 and $200 for each kilowatt of generating capacity. Because of the difficulty in fitting scrubbers into old power plants not originally designed to accommodate them, a scrubber for an existing plant costs between 10 and 40 percent more than a scrubber for a new plant—assuming that the installation goes smoothly. Regardless of the age of the plant, the cost of operating the scrubber is high. Scrubbers create huge amounts of sludge waste that must be put in holding ponds and landfills, constituting an environ-

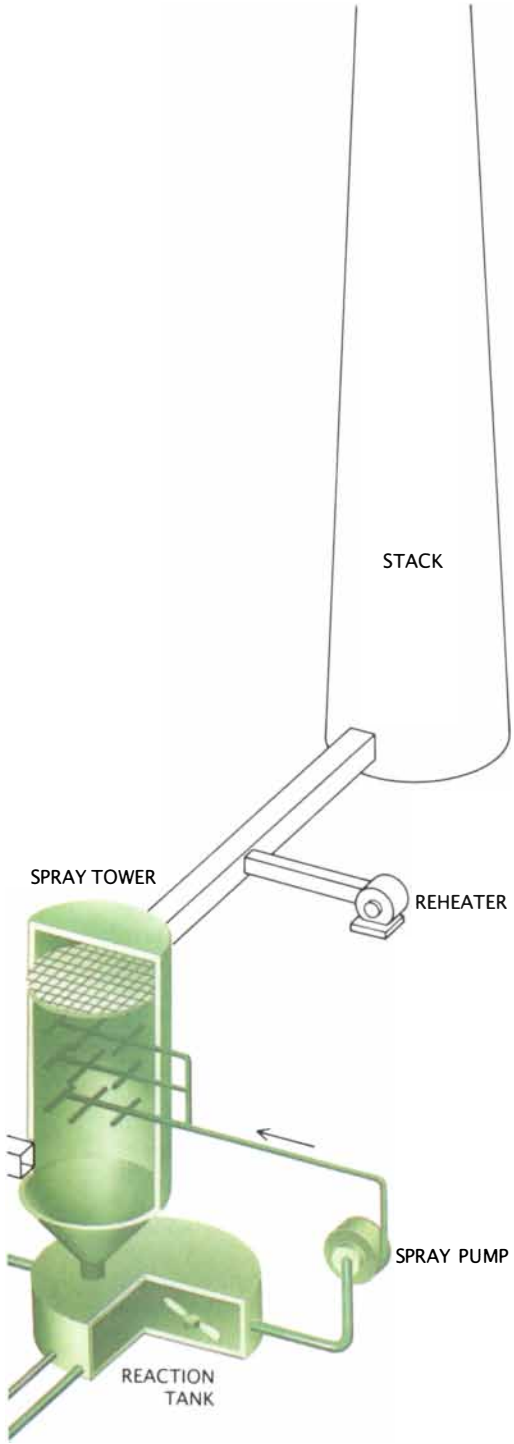
COOL WATER DEMONSTRATION PLANT of the Southern California Edison Company converts 1,000 tons of coal daily into a clean-burning gas whose combustion products drive a gas-turbine generator. The "waste heat" in the exhaust from the gas turbine then produces steam that drives a steam-turbine generator. The aerial photograph shows the plant's two coal-storage silos (*middle*); the gasifier facility, gas-cooling system and electricity-generating equipment extend from the silos to the lower right.

© 1987 SCIENTIFIC AMERICAN, INC



TYPICAL CONVENTIONAL PLANT burns pulverized coal in a boiler to produce steam, which then drives a set of turbines connected to a generator. The steam is then condensed and recycled through the boiler. Much of the other equipment at the plant is intended to remove pollutants from the flue gas, or boiler exhaust, before it is released into the air. First, an electrostatic precipitator collects dust particles from the flue gas by channeling the gas between charged electrodes and grounded plates. The particles acquire a charge from the electrodes as they pass by and are attracted to the plates. Next, a flue-gas desulfurization system, known as a wet scrubber (*green*), extracts most of the oxides of sulfur from the gas. This is generally achieved by

© 1987 SCIENTIFIC AMERICAN, INC



bringing the gas in contact with a slurry of limestone (calcium carbonate) and water. The sulfur oxides react with the slurry to produce a sludge consisting of either calcium sulfite or calcium sulfate (which can be sold as building material).

mental side effect in itself. A 1,000-megawatt power plant that burns coal with a sulfur content of 3 percent, for example, produces enough sludge in one year to cover about a square mile to a depth of one foot.

Wet-scrubbing systems also require large amounts of water (about 1,000 gallons per minute for a 1,000-megawatt plant), and they often suffer from plugging and fouling of equipment and corrosion of ductwork. These problems add to operating costs and reduce the overall reliability of the system. Finally, scrubbers extract a penalty of between 3 and 8 percent of the power plant's energy output simply to run pumps and fans and to reheat the flue gas in order to prevent corrosive condensation in the chimney.

The widespread implementation of scrubber technology in the U.S. has not been easy or cheap. The reliability of early units was considerably less than that of the power plant as a whole; hence they required components that were either redundant or had been engineered with large margins of tolerance. Some of the difficulties associated with the installation and operation of scrubbers can be attributed to the fact that the technology was prematurely brought up to commercial scale. Only now, after a quarter-century of experience, have scrubbers approached an acceptable level of reliability.

Coal-fired power plants have become more expensive to build not only because of emission-control systems but also because construction costs themselves have skyrocketed. Even after adjusting for inflation, new coal-fired generating capacity is three times as expensive today as it was in 1970. Over the past 15 years the economies of scale that are to be gained from constructing large plants have been offset by large cost increases. In part the increases reflect the high cost of financing prolonged construction delays.

The experience of Japanese utility companies demonstrates how big a difference construction delays can make. Japanese utilities typically deal more expeditiously than their American counterparts with the various regulatory and financial issues that often delay large construction projects. Whereas the Japanese are able to bring power plants on line in from 35 to 40 months, American companies generally require from 50 to 60 months for plants of similar design. Consequently the cost of a single new large generating unit is comparable to the total assets of many U.S. electric utilities.

Utilities are therefore seeking to reduce the costs by turning to smaller, modular generating units, which can be transported to the sites of existing power plants and rapidly installed there to keep pace with the growth of demand. Because such units can be brought on line quickly, their capital cost can be recovered faster—even if the rate base remains constant. Installing new units only when new capacity is needed can result in a net saving of as much as $200 per kilowatt, in spite of the fact that smaller units sacrifice economies of scale.

As an alternative to building new generating capacity, utilities have been refurbishing old plants to improve their performance and extend their lifetime. This strategy is certainly less costly than building a new plant. The trend is reinforced by the fact that plants now completing three decades of service are not obsolete. In some cases they may even be more efficient than new plants, since they are not burdened by scrubbers. Indeed, older plants are becoming a larger fraction of our total electricity-generating capacity. In 1970 only 2 percent of U.S. generating capacity was more than three decades old. By the end of the century 30 years will be the average age of a coal-fired power plant.

Utilities are also looking at ways to lower operating costs. Early warning of the deterioration of critical system components is essential to prevent the loss of generating capacity. Hence the continuous monitoring for incipient failure is becoming an important part of maintenance procedures. Such monitoring keeps track of the natural processes of wear, corrosion and erosion, making it possible for plant operators to institute corrective measures in time to avoid forced electricity outages. The value of such efforts can be appreciated by recognizing the fact that an idled 1,000-megawatt coal-fired plant may cost the utility as much as $1 million for every day it is out of service, largely because the electricity must be bought from other, more costly sources.

Escalating costs for the transportation and handling of coal as well as for ash disposal have also made the quality of coal (determined from its content of moisture, sulfur and other minerals) an important consideration in improving plant performance. Although low-quality coal may well be

© 1987 SCIENTIFIC AMERICAN, INC

cheaper per ton than high-quality coal, more of it must be burned to generate the same amount of electricity. The cost of shipping a greater volume of low-quality coal may offset the cheaper price. Low-quality coal typically generates more waste than higher-quality coal, leading to higher costs for waste disposal. Finally, the composition of low-quality coal is more variable than that of the better grades, making it difficult to "tune" a plant's systems to achieve the greatest possible efficiency: the systems have to be adjusted to cope with the worst quality of coal expected in a load.

Existing plants can improve the coal or at least maintain a uniform quality by removing some of the impurities, such as sulfur-containing minerals, before the coal is burned. This is done at facilities that crush the "dirty" coal and then separate the coal from the mineral impurities by virtue of differences in specific gravity or other physical properties.

In spite of such measures for improving the performance of existing coal-fired plants, the U.S. may still need 150,000 megawatts or more of additional generating capacity by the end of the century, if demand for electricity grows at the projected annual rate of 2.3 percent. To maintain the competitiveness of coal in such an expanding energy market, utilities will have to adopt coal-burning technologies that are more cost-effective than conventional ones in three key areas: pollution control, construction and performance.

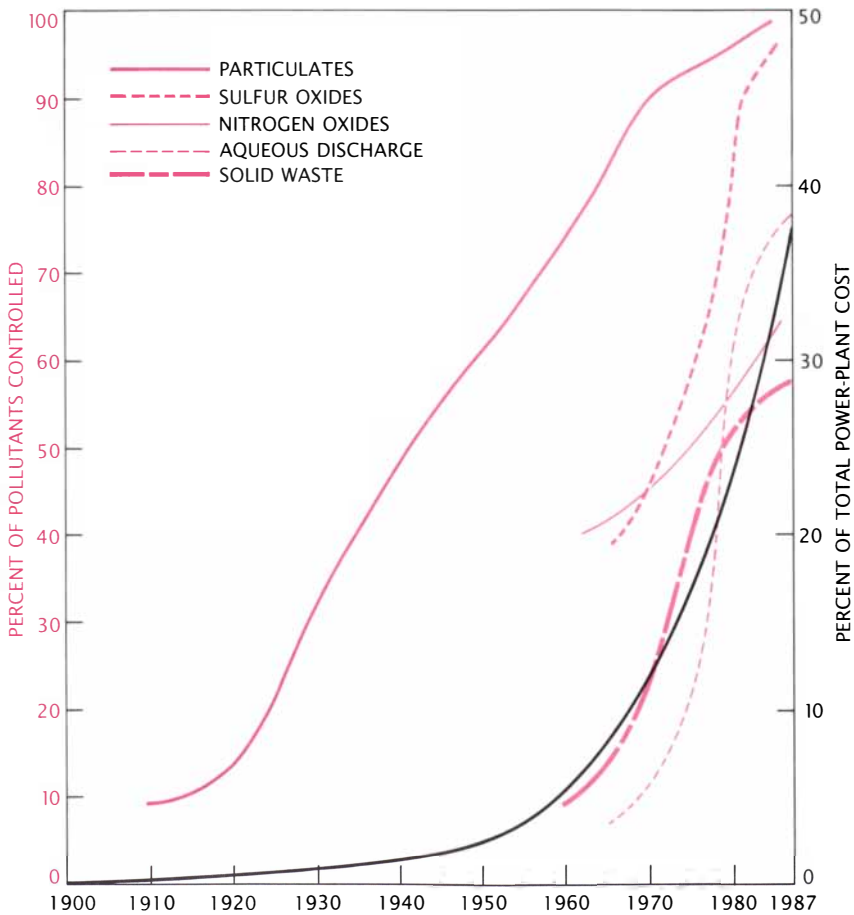
There are a dozen or more advanced technologies under development for burning coal cleanly and efficiently. Two of the more promising are fluidized-bed combustion (FBC) and coal gasification. In a typical FBC coal boiler, crushed coal is "fluidized" along with bits of limestone by supporting the particles on a strong rising current of air. The fluidized particles act as if they were in a boiling liquid, that is, they mix turbulently, ensuring a very efficient combustion process. Because boiler tubes are in direct contact with the fluidized particles, an FBC boiler can absorb much of the heat by conduction, which is more efficient than the radiant and convective heat transfer a conventional boiler relies on.

The greater surface area for heat transfer offered by the tubes in the fluidized bed also enables an FBC boiler to operate at lower temperatures than a conventional pulverized-coal boiler does, thereby minimizing the formation of oxides of nitrogen. (Whereas the temperature in a conventional boiler may reach 3,000 degrees Fahrenheit, in an FBC boiler it generally ranges between 1,450 and 1,600 degrees.) Moreover, the limestone mixed with the coal captures 90 percent or more of the sulfur released from the coal during combustion, because the lower operating temperature promotes the reaction between limestone and sulfur that produces calcium sulfite or sulfate. FBC tackles the pollutants produced by burning coal where they are formed: in the furnace.

FBC boiler design and operation are also less sensitive to variations in coal quality. The large amounts of melted ash that result when low-quality coal is burned in a pulverized-coal boiler often foul heat-exchange surfaces, reducing the boiler's efficiency and reliability. In an FBC boiler, however, coal is burned at temperatures below the melting point of ash, avoiding the problem. Because an FBC plant operator can afford to buy lower quality coal, he can in some instances substantially reduce operating costs.

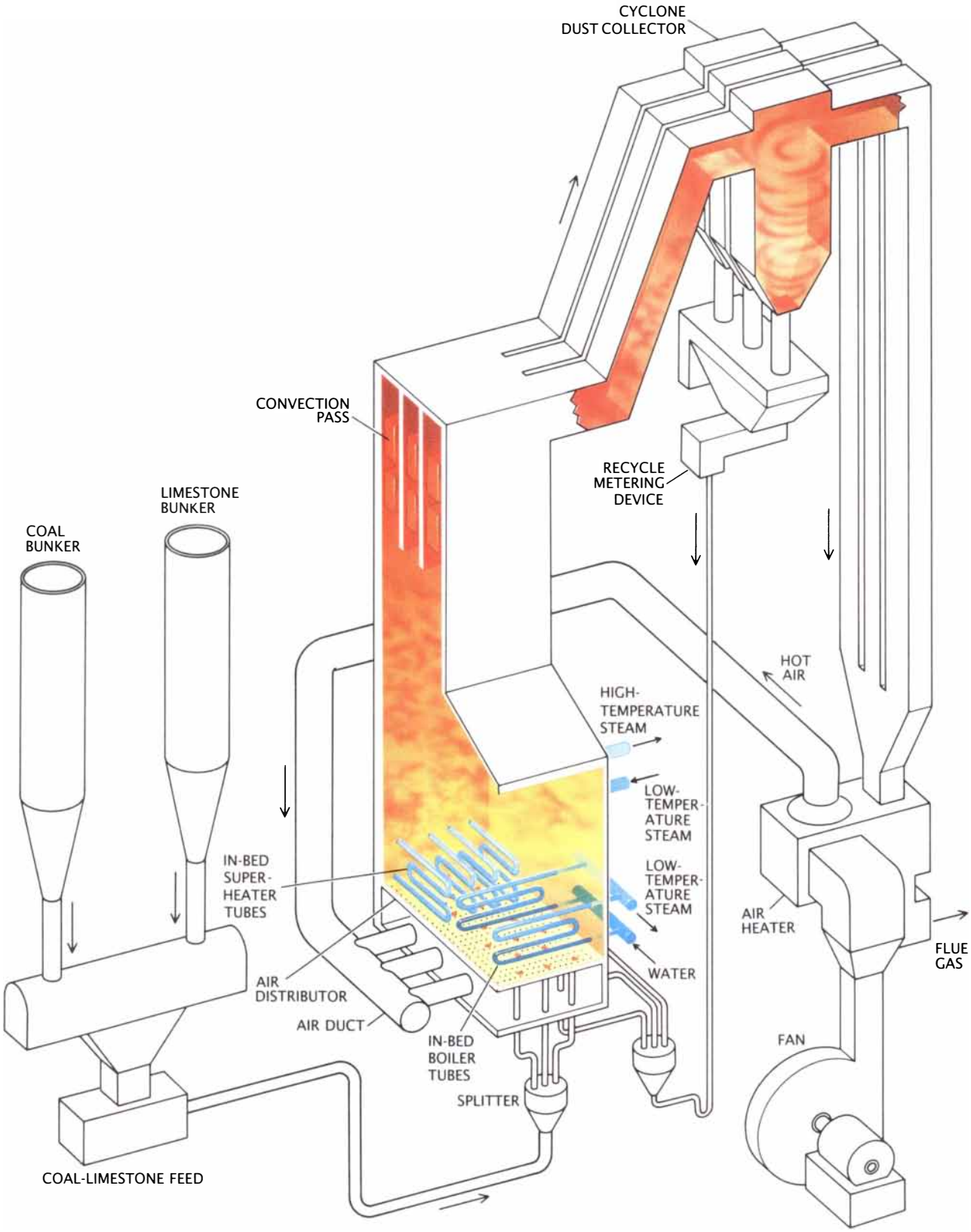
FBC technology also lends itself to the design of small boilers that can be prefabricated as modular units. It is estimated that a plant built from compact, pressurized FBC boilers could require between 10 and 20 percent less investment capital than conventional plants having the same total generating capacity. The savings result from a reduced construction time and better capacity to expand the plant in response to an uncertain growth in the demand for electricity; planning is also made easier by the knowledge that these compact units can be installed quickly whenever the need arises.

An FBC boiler could also be added to an existing plant for a rap-



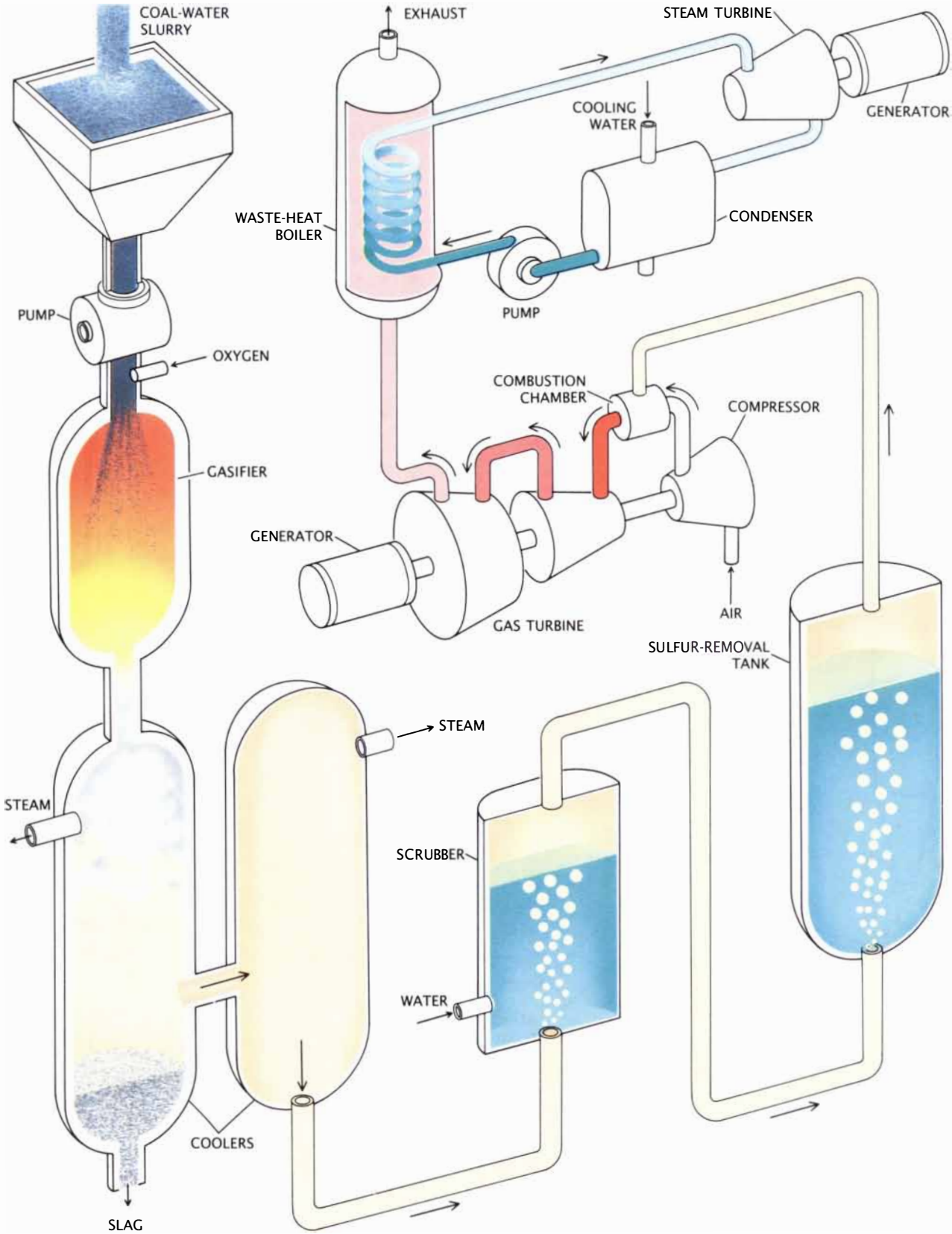
COST OF CONTROLLING power-plant emissions and waste products as a percentage of the total plant cost (*black*) has increased in step with the percentage of the emissions or waste brought under control (*color*). The pollution-control systems mandated by regulatory agencies now account for between 30 and 40 percent of the total cost of a plant.

© 1987 SCIENTIFIC AMERICAN, INC



FLUIDIZED-BED COMBUSTION (FBC) of coal reduces the need for adjunct emission-control systems. In an FBC boiler a bed of crushed coal and limestone is suspended by a blast of air that fluidizes, or agitates, the bed particles so that they behave as if they were in a boiling liquid. The turbulent mixing ensures that the coal is thoroughly burned and that the limestone reacts with combustion gases and thereby captures 90 percent or more of the sulfur oxides. Because boiler tubes are in direct contact with the fluidized bed, steam can be produced more efficiently in FBC boilers than in conventional pulverized-coal boilers. An FBC boiler also operates at a lower temperature, avoiding the melting of ash and reducing the formation of nitrogen oxides.

© 1987 SCIENTIFIC AMERICAN, INC



COAL GASIFICATION can be accomplished by heating a coal-water slurry in an oxygen atmosphere. The process yields a fuel gas composed mainly of carbon monoxide and hydrogen. After the gas has been cooled, cleansed of solid particles and rid of sulfur it can be burned to drive gas turbines and then produce steam for a steam turbine, an arrangement called a combined cycle. An integrated gasification/combined-cycle plant emits fewer pollutants into the air than conventional coal-fired plants do.

© 1987 SCIENTIFIC AMERICAN, INC

id increase in generating capacity. Indeed, the Northern States Power Company has converted an existing coal-fired boiler in a Minnesota plant into an FBC boiler. The objective is to increase generating capacity by 40 percent, increase fuel flexibility (the boiler can burn even municipal refuse), control emissions more thoroughly and extend the plant's operating life to at least 50 years.

Over the past 15 years the technology of power plants that rely solely on FBC boilers has advanced from small test and pilot-plant facilities to large demonstration plants. The Tennessee Valley Authority, the Duke Power Company and the Commonwealth of Kentucky are building such a plant with a generating capacity of 160 megawatts; the Colorado-Ute Electric Association, Inc., has begun operating a 110-megawatt FBC-based generating unit. If they are successful, these two projects along with the Northern States Power Company's project—a combined private-sector commitment of nearly $400 million—will reduce the economic risk entailed in the application of FBC boilers for the utility industry.

An alternative technology, which—in less sophisticated form—was actually commonplace in the mid-19th century, produces clean-burning gas from coal [*see illustration on preceding page*]. The gas is suitable for lighting and heating purposes and in fact had an important role in the U.S. until it was displaced by natural gas just before World War II.

Utilities were initially interested in coal-gasification processes because they can produce a fuel for boilers that burns cleanly and thereby eliminates the need for scrubbing. It is now clear that coal gasification has a more significant advantage: turbines can be driven directly by the hot gases produced by the combustion of coal gas. This in turn makes it possible for the "waste" heat of the combustion products leaving the gas turbine to be captured and harnessed to produce steam for a steam turbine. Such an arrangement of coupled gas and steam turbines is called a combined cycle and is one of the most efficient means available to generate electric power.

Coal gas from which sulfur and particulate matter have been removed is an excellent gas-turbine fuel and, like natural gas, can be burned with very low emissions. The combined cycle's high efficiency compensates for the unavoidable efficiency losses that result from converting coal into a gas. Moreover, water consumption is lower than in a conventional coal plant because two-thirds of the power is produced by the gas-turbine system, which needs no cooling water to condense steam.

The viability of an integrated gasification/combined-cycle (IGCC) power plant has been demonstrated at the Cool Water site of the Southern California Edison Company. The Cool Water plant, which began operating in May, 1984, generates about 100 megawatts of electricity from a wide range of coal types; its emissions are comparable to those of a nearby natural-gas power plant. Sulfur oxides in its flue gas are kept well below Federal standards by means of an adjunct sulfur-removal system that extracts essentially all the sulfur originally contained in the coal to yield pure sulfur with a significant market value. Nitrogen oxide formation is suppressed by adding moisture to the gas before it is burned, so that the combustion temperature of the gas is lowered. Moreover, the coal residues left in the gasifier are melted, yielding an inert, vitreous material when cooled that meets California solid-waste standards.

In addition to its advantages in pollution control and efficiency, an IGCC plant can be constructed in phases, allowing generating capacity to be built up in increments. Such flexibility in construction minimizes the risk in capital investment associated with an uncertain growth of the electric load. The first increment of generating capacity, for example, might be provided by gas turbines, which, taking advantage of current low fuel prices, might be driven not by coal gas but by oil or natural gas. Then, as the demand for electricity grows, the addition of a waste-heat boiler and steam turbine would increase not only the capacity of the plant but also its efficiency. When oil and gas prices as well as load requirements make it economical to do so, a coal-gasification facility could then be constructed at the plant site.

The role of coal-fired power plants is a crucial question in the public-policy debate over the stewardship of natural resources, the protection of the environment, the provision of energy and the growth of the economy. Yet these considerations need not conflict: a look at the emerging technological options shows that emission controls and cost reductions can in fact complement one another. This principle was recognized in a joint U.S.-Canadian report on acid rain issued last year. Acting on the report's recommendations, Congress is currently considering the establishment of a major national initiative to demonstrate and deploy clean coal-burning technology. This initiative, which would augment private-sector resources with Federal investments, is intended to ensure that the full range of coal-processing technologies—including FBC boilers and coal gasifiers—will be available for widespread commercial application in the 1990's. Even with such power-plant technologies in the offing, however, the growing demand for electricity could not be met without a collection of concurrent measures that emphasize conservation, end-use management and increased productivity from existing power plants.

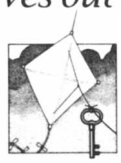
Continuing economic and environmental considerations will eventually lead to technological developments extending well beyond those we have discussed here. Ultimately coal-fired power plants could evolve into fully integrated "resource refineries." Such facilities would adapt themselves to the local fossil-fuel and resource base to produce a mixture of electricity, heat, fuels and marketable products for the local economy. In addition to FBC boilers and coal-gasification facilities, an entirely new spectrum of technologies—linked by electronic diagnostic and control systems—would be incorporated at such plants to recover most of the by-products of coal processing.

There is no shortage of opportunities to improve both the economic and the environmental aspects of coal-based electric-power production. The timely exploitation of these opportunities, however, depends on whether the Government can enact broad and balanced energy and environmental policies that give the electric-utility industry the necessary incentives. Care must be taken that innovative coal-processing technologies are developed and implemented rationally, with the cooperation of utilities—in contrast to the way scrubber technology was introduced. This could be ensured if the associated costs and risks are minimized through well-planned design, testing and refinement of small-scale prototypes and demonstration plants before commercial-scale systems are finally introduced.

© 1987 SCIENTIFIC AMERICAN, INC

THE AMATEUR SCIENTIST

Sticky threadlike substances that tend to draw themselves out into bead arrays



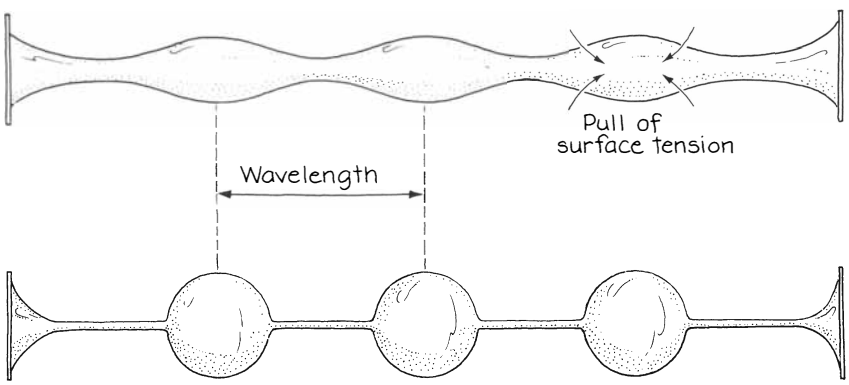
by Jearl Walker

The tendency of a threadlike liquid to draw much of itself into beads has intrigued many investigators. The beads can be found on sticky strung-out liquids such as the circular strands of certain cobwebs and a bit of saliva drawn out between the thumb and forefinger. C. V. Boys, author of the classic *Soap Bubbles and the Forces Which Mould Them,* demonstrated another form in a lecture at the London Institution many years ago. He dipped a straw into castor oil and coated a quartz fiber thinly with the oil, which quickly pulled itself into beads less than .01 millimeter in diameter.

Lord Rayleigh examined beading in a set of papers published in 1879 and 1892. His work was based on earlier experiments done by the Belgian physicist Joseph A. F. Plateau, who is still renowned for his pioneering study of thin films. When a liquid is drawn into a thread, surface tension minimizes the surface area by pulling the thread into a cylindrical shape. At the same time, however, waves playing along the thread work to distort it into a periodic series of narrow and wide regions. The waves are unavoidable, arising from the motion when the thread is drawn out, from any slight shaking of the mount and from motion of the air.

Plateau demonstrated that the wavelength of a wave determines whether the thread resumes a cylindrical shape after the wave has passed. If the wavelength is less than the circumference of the thread, the distortion produced by the wave increases the overall surface area. The area of the narrowed regions decreases but that of the widened regions increases much more. As soon as the wave has passed, surface tension pulls the thread back into a cylinder to minimize the surface area.

If the wavelength is greater than the circumference, the distortion decreases the overall surface area. The area of the narrowed regions decreases more than that of the widened regions increases. Then surface tension enhances the distortion, pulling the widened regions into beads and the narrowed regions into a thread thinner than the original one. The spacing of the beads is equal to the wavelength of the wave. A weak thread may rupture, freeing the beads. If a thread is strong, additional waves may create a second set of smaller beads mixed in with the first set—all of them strung together by a thread that may now be virtually imperceptible.



Evolution of beads on a liquid thread

Rayleigh found that the initial liquid thread is most unstable when a passing wave has a certain critical wavelength. A wave of that length generates beads faster than waves of other lengths and therefore dominates the formation of beads. If the viscosity of the liquid can be disregarded, the critical wavelength is approximately one and a half times the circumference of the thread. Waves of about that size leave a number of beads separated by a distance roughly equal to the initial circumference of the thread. When the viscosity is high, as in a thread of molten glass, the critical wavelength is much longer. Beads are fewer, and they are farther apart.

A commoner example of a strung-out liquid is a soap film suspended between two mounts. Jurjen K. van Deen of The Hague has studied several of the stability limits in such an arrangement. (I described some of his equipment in this department last month.) For his experiments on stability he prepared two rings, each of them 50 millimeters in diameter, by soldering copper wire. The rings are mounted horizontally, one above the other, on a stand.

To form a bridging film van Deen blows a soap bubble between the rings and makes sure that it anchors on them. Since the solution gradually drains along the film, fresh solution is provided through a tube near the top of the bubble. An extension of the wire forming the top ring fits loosely into the tube, guiding the solution to the bubble. As liquid drains to the bottom of the bubble, it tends to collect in a drop that makes the bubble oscillate when it detaches. Van Deen eliminates the oscillations by placing a length of copper wire just below the bubble. Because of the wire, the drop detaches gently.

The bubble forms a sphere because surface tension pulls inward, tending to minimize the surface area. Air pressure inside the bubble counters the inward pull. (Actually the cause is the excess of the inner air pressure over the atmospheric pressure.) The bubble is stable when its air pressure is equal to $4\pi S/r$, where S is the film's surface tension and r is the radius of the bubble.

In one set of experiments van Deen reduces the bubble's air pressure with a vacuum cleaner, which sucks air out by means of a tube stuck

© 1987 SCIENTIFIC AMERICAN, INC

through the top ring. A clamp on the tube controls the rate at which air is removed. As the air pressure drops, the bubble begins to change shape. When the pressure reaches $2\pi S/r$, the sides of the bubble form a cylinder. This shape requires less pressure than a sphere because the net curvature of the cylinder is less than that of a sphere. (A cylinder does not curve along its length.)

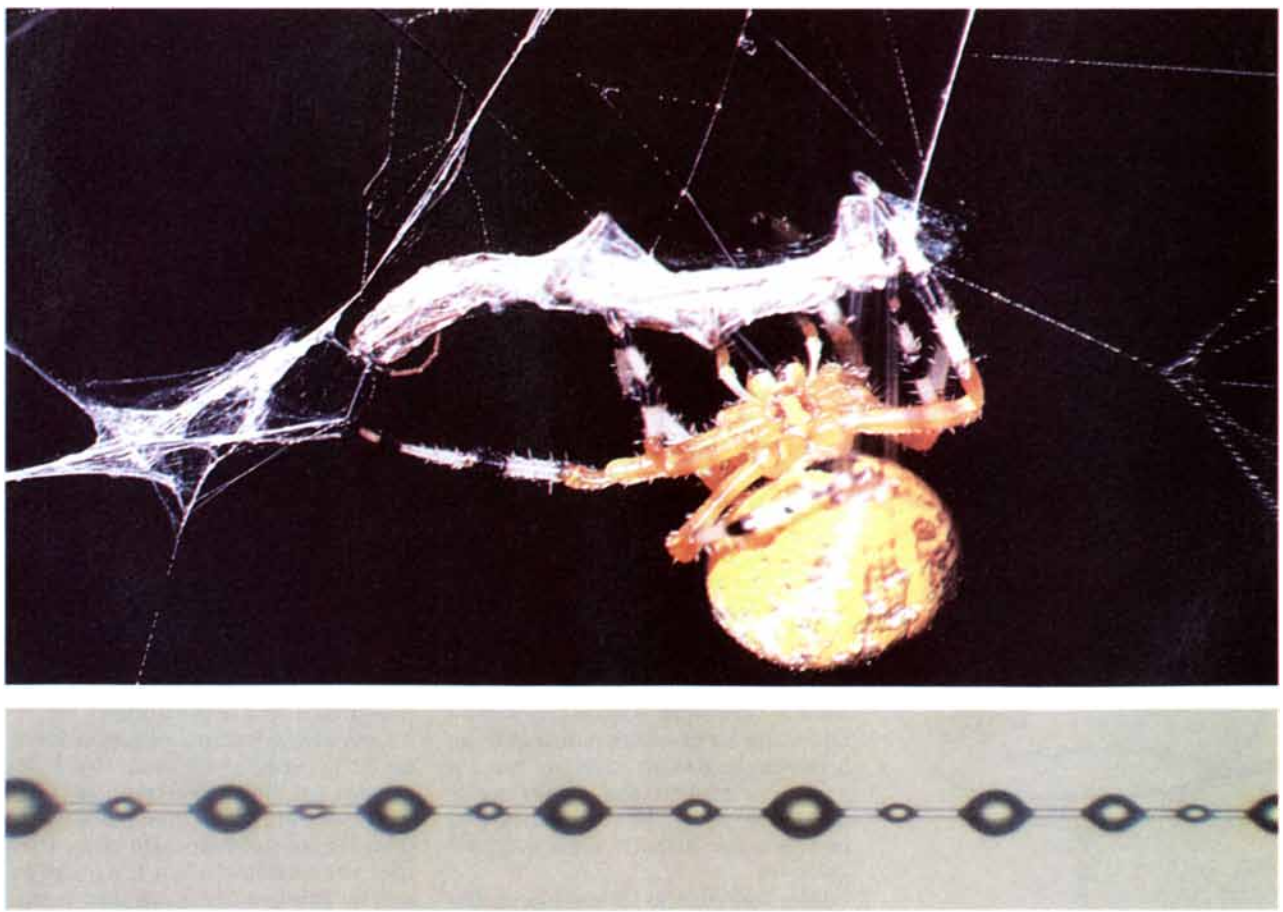
A cylindrical bubble is peculiar in that its existence depends on the ratio of L, the separation between the rings, to d, the diameter of the rings. If the ratio exceeds π, the bubble becomes unstable and collapses to form two smaller bubbles, one on each ring. This limit on stability is less surprising if it is recast in the following form: a cylindrical bubble can exist if L is less than the circumference of the cylinder.

In this form the limit of stability is related to the beading on a liquid thread. Remember that the waves responsible for beading have a wavelength longer than the circumference of the thread. A cylindrical bubble also develops waves from chance disturbances, but their wavelength cannot be longer than L. Hence if L is kept smaller than the circumference of the cylinder, the waves are too short to make the cylinder collapse. Every time a wave passes along the cylinder the overall surface area increases. Surface tension immediately pulls the bubble back into a cylinder. If L exceeds the circumference, the waves can be at a wavelength that exceeds the circumference. Such a wave decreases the overall surface area of the bubble. Surface tension tends to distort the cylinder. As a result the center of the bubble collapses, leaving a spherical bubble on each ring.

In another experiment van Deen anchors a spherical bubble to the rings and pricks the two sections enclosed by them. The ruptures open the bubble to the atmosphere, whereupon the internal pressure drops to zero. This change causes the curvature of the remaining soap film to become zero also. How can the film stretch between two rings without curvature?

The explanation is that the bubble forms a catenoid: a surface generated when a catenary curve is rotated about an axis. (A catenary curve is formed when a stringlike object is suspended between two fixed points.) The catenoid's surface constitutes the minimum area for a film stretched between two rings when there is no pressure inside the structure. The surface also has a net curvature of zero. Although the surface is in fact curved, any point on it is part of a concave curvature in one direction and a convex curvature in a perpendicular direction, so that the net curvature is zero.

This configuration also sets a limit on the film's stability that involves the ratio of L to d. The catenary shape can exist only if L/d is less than .663. If the limit is exceeded, the waist of the catenoid suddenly collapses, leaving a flat film on each ring. The reason is that when L/d exceeds .663, the flat films have a combined area smaller than the area of the catenoid. (Actually the combined area of the flat films is smaller than that of the catenoid when L/d exceeds .528. The catenoid can be stable between .528



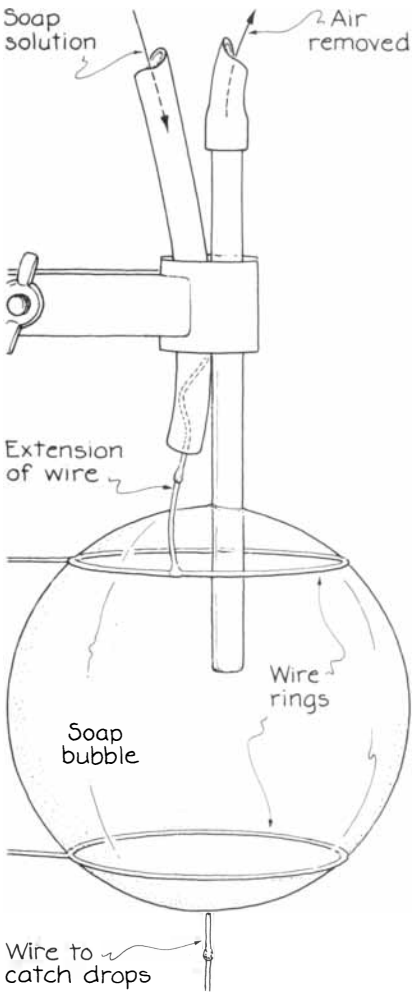
Spider on a beaded web and a micrograph of a beaded thread

© 1987 SCIENTIFIC AMERICAN, INC

and .663, however, because its surface area would have to increase momentarily if it could transform to two flat films. Unless the waves on a soap film are vigorous, surface tension causes the catenoid to re-form after the passage of each wave. It fails to do so when L/d exceeds .663. Then even a feeble wave gives rise to an immediate collapse.)

Van Deen discovered that maintaining a stable catenoid with L/d greater than .5 was difficult. The catenoid nearly always collapsed because of waves from various disturbances. If you experiment with a catenoid film, isolate it from vibrations and air currents.

In the early 1970's M. A. Erle, R. D. Gillette and Derek C. Dyson of Rice University also studied the stable shapes of soap films, but in their arrangement the film formed a bridge between two coaxial, parallel disks. The workers dealt with films that were symmetrical around the central axis and with symmetrical disturbances to the films. In their experiments the film was anchored to protruding rims on the disks. Air was removed from or added to the film through a small hole in one of the disks. When the film was inflated, it eventually reached a limit of stability, a condition in which its sides bulged outward. When more air was pumped in, the film broke free.



Jurjen K. van Deen's soap-bubble setup

When air was removed from the film, the sides moved inward until it reached another limit of stability. Then the center collapsed, leaving a hemispherical film on each disk. The shape of the film just before collapse again depends on L/d. The limiting shape is an inward bulge when the ratio is less than π and an outward bulge when it is greater than π.

Suppose that L/d is less than π. Visualize the change in the film's shape as air is removed. Begin with the film at the limiting shape of an outward bulge. As the volume of air in the film decreases, the structure shrinks and its curvature increases, raising the inner air pressure. For any given inner volume of air the film has a unique shape determined by the interplay of the net curvature and the air pressure. When more air is taken out, the curvature begins to decrease and the air pressure drops. When the curvature reaches zero, the side of the film is a catenary.

This configuration is called a type-1 catenary. A type-2 catenary can also have a net curvature of zero, and so be compatible with an air pressure of zero inside a film. The mathematical functions for both catenaries are similar, but the type-2 catenary has a narrower waist and a larger surface area. When the film is on rings, surface tension prevents the formation of a type-2 catenary; instead it decreases the surface area by pulling the film into a type-1 catenary.

A type-2 catenary is possible when the soap film is mounted on disks and the volume of air in the film is reduced. Beginning with the film in a type-1 catenary, imagine the change in its shape when more air is removed. Its sides begin to move inward as the inner pressure drops below the external air pressure, and the net curvature is outward. After a minimum air pressure is reached the pressure begins to increase until it again reaches zero and the net curvature is also zero. The curve of the film between the disks is then a type-2 catenary.

Erle, Gillette and Dyson found that a film with a type-2 catenary is possible if L/d is between .47 and .663. At any value less than the lower limit the film always collapses into two hemispherical shapes before the internal pressure reaches zero for a second time. If L/d is, say, .5 and the soap solution is long-lasting, the film can survive as a type-2 catenary for more than an hour.

The research team also investigated the stability of greatly inflated films. Theoretically, if the film is symmetrical with respect to the central axis and if only symmetrical disturbances are considered, the film can bulge outward quite significantly before it becomes unstable. The calculated limit is reached when the surface of the bridge is perpendicular to the back of each disk.

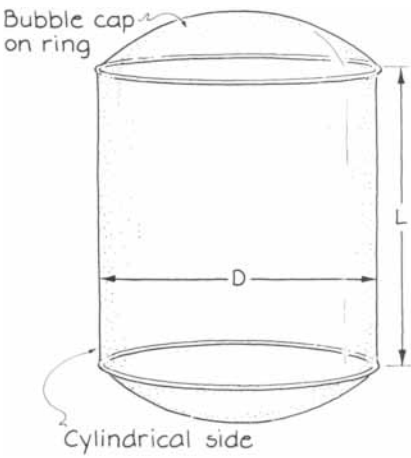
Recently M. R. Russo and Paul H. Steen of Cornell University considerably reduced this theoretical limit by examining asymmetrical disturbances. They prepared a silicone-oil bridge that stretched between two brass plates. The bridge was surrounded by a mixture of *n*-propanol and water that had the same density as the oil. The match of densities effectively removed the distortion normally caused by gravity because the buoyancy on the bridge was equal to the weight of the bridge.

The liquids were put in a Plexiglas cell to eliminate evaporation and consequent changes in density. The oil was added to or removed from the bridge by means of a flexible tube inserted through a small hole in one of the plates. A syringe at the other end of the tube facilitated the change in the volume of the oil.

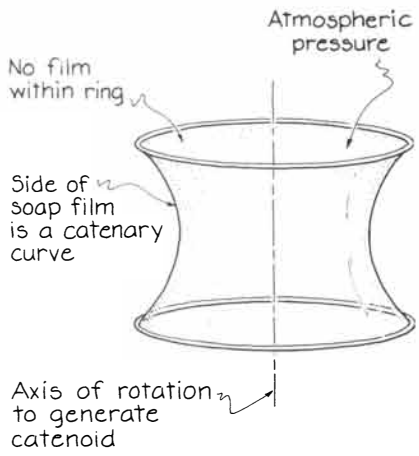
Beginning with a cylindrical oil bridge, Russo and Steen decreased the distance between the plates by turning a micrometer drive. Because the volume of oil remained constant, the bridge billowed outward. Each plate's beveled edge and coating of paraffin wax helped to anchor the oil on the plate. The bridge remained symmetrical until its surface at the plates was tangent to the plane of the plates. When the distance between the plates was further decreased, increasing the bridge's outward bulge, the bridge suddenly became distorted on one side. The buckling reduced the surface area of the bridge.

Cylindrical bridges of liquid have been the subject of study by J. M. Haynes of the University of Bristol, who considered their stability as their length increases. He knew that they are unstable when L/d exceeds π. The bridges break up into hemispheres on the plates to which they are anchored. Haynes also discov-

© 1987 SCIENTIFIC AMERICAN, INC



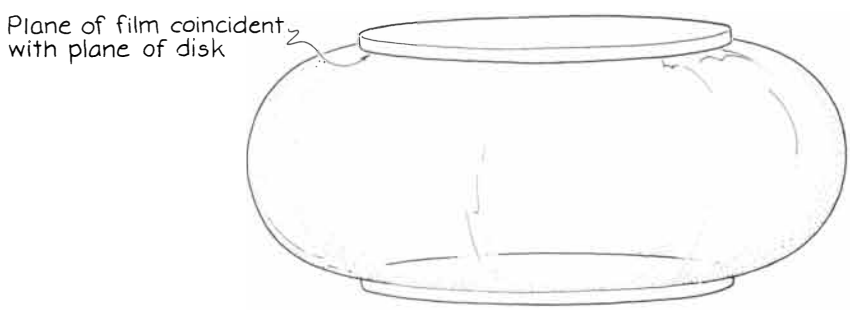
A cylindrical bubble on two rings



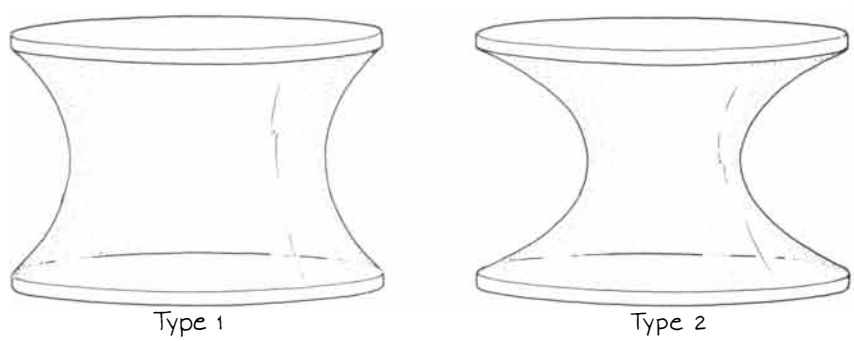
The catenoid surface

ered that they show surprising stability at smaller ratios. When L/d exceeds 9/4, the surface area of a cylindrical liquid bridge is actually more than the total surface area of the hemispheres. If the ratio is less than π, however, the bridge does not collapse to form hemispheres because to do so it would momentarily have to increase its surface area. That barrier disappears when L/d reaches π.

Several readers have commented on what I said in June about the Minotaur Cube. They found two solutions instead of one. The first letters to arrive were from John Stewart and Jim Rostirolla of Bellevue Community College in Bellevue, Wash., Leonard Gordon of Chico, Calif., and Michael Keller of Ellicott City, Md. Keller found the first solution "by hand" and then was surprised when his computer turned up the second one. The solutions have identical plays for pieces *A* and *B*. The problem as I gave it was to employ the six pieces of the puzzle—three quadcubes and three pentacubes—to form a cube with three units on a side.



The outward-bulge limit



Types of catenaries

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 0036-8733/87. $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. 11737–0000. 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

© 1987 SCIENTIFIC AMERICAN, INC

COMPUTER RECREATIONS

Diverse personalities search for social equilibrium at a computer party

by A. K. Dewdney

I recently gave a party I was not able to attend in person. My eight guests had quite distinct personalities and occupations; for example, there was Walter (not his real name) the weight lifter. Wally had conceived a hopeless passion for another guest, Princess Penelope, a refined and sensitive woman of aristocratic pretensions. At the party Wally continually edged closer to Princess Penelope, while she just as continually moved away from him. In quantitative terms Wally would have liked to spend the entire evening just three feet away from Penelope; to have been any closer would have been socially unacceptable. For her part, Penelope was not comfortable unless 15 feet separated her from her admirer. But if Wally strayed any farther, Penelope would begin some edging of her own. Perhaps she enjoyed remaining in Wally's view.

It will surprise few readers to learn that the party was held in my computer. The guests were represented on my display screen as typographic characters: Wally was a *W* and Penelope was a *P,* for instance. As the party progressed, the eight guests drifted about the room in a seemingly endless search for social equilibrium. Occasional clusters of closely interacting people would form only to be dispelled by the arrival of a new individual who upset the delicate balance in some subtle way. Each guest had an ideal distance from each of the others. He or she would always move in a direction to minimize unhappiness. In the context of my admittedly quantitative (and stereotyped) party, the unhappiness of a given guest was measured by the sum of the differences between the ideal and the actual distances to all the other guests.

NAME	OCCUPATION
ARTHUR	ARTIST
BERNIE	BUSINESSMAN
DENNIS	DENTIST
MILLIE	MODEL
PENELOPE	PRINCESS
SUSAN	STOCKBROKER
VIOLA	VIOLINIST
WALLY	WEIGHTLIFTER

Guest list for a computer party

The guests were confined to a single rectangular room dominated by a table spread with tempting refreshments. Each person had an ideal distance from the table as well as from the others; dieters preferred a distance of five feet, whereas others were not completely happy unless they were just a foot from the table. This is not to say that the average, nondieting guest would simply move toward the table. For example, Arty the artist might have wanted to sample some digital dip but Bernie the businessman (whom Arty disliked) blocked the way, expounding on the principles of vertical marketing.

The program that produced the party is called PARTY PLANNER by its originator, Richard Goldstein, a computer artist and game designer who lives in Los Angeles, Calif. Rich Gold, as he prefers to be called, is well known as the designer of "The Little Computer People," a game that produces a cross-sectional image of a house in which a mannequin (the little computer person) goes about his daily routine.

Gold calls his new recreation PARTY PLANNER with tongue in cheek; he maintains that by coding the likely relations between guests invited to a real party one can determine possible outcomes of the mix in terms of social dynamics. One might even discover the optimal location for the refreshment table. Perhaps a proper evaluation can only be left to those readers who both write their own version of the game and attempt to apply it to real life. The real guests, it would seem, may at least chuckle at their own antics on the screen.

Gold's computer party can be held in a room digitized into 600 squares in a 20-by-30 array. Each guest occupies a single square of the array. PARTY PLANNER proceeds by considering each person in turn; the program moves a guest to each of the eight neighboring squares. In each location the program calculates the total unhappiness of the guest. The neighboring square in which unhappiness is at a minimum becomes the person's new location.

Organizing a party requires only a modicum of programming experience. The 20-by-30 array will be called *room.* Since the entries of the array are alphabetic characters, *room* must be declared as a character array in the program. The walls and refreshment table are designated by fixed characters [*see illustration on opposite page*].

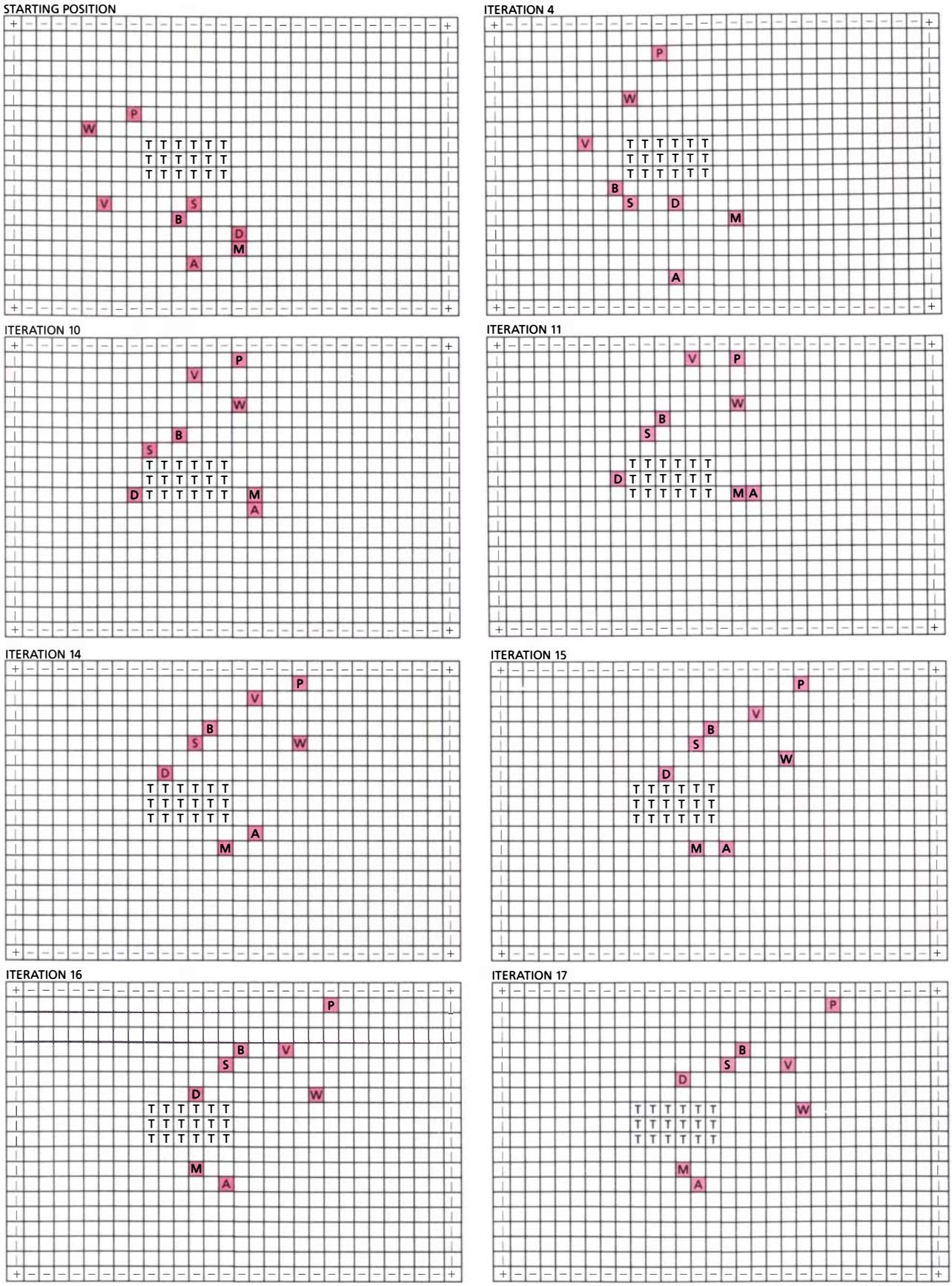
PARTY PLANNER has a fairly simple structure that can be described on two levels, from the outside in. First there is a loop within which all the guests are moved about the room endlessly, or at least until the user of the program presses the space bar. In algorithmic terms the outer level can be summarized as follows:

```
repeat
    for i ← 1 to 8
        move the ith guest
    display room
until space bar
```

Inside the repeat loop is a second loop in which the eight people are all shifted to new locations. I shall expand below on the mysteries of moving the *i*th guest. A display of the room requires a double loop that scans the array *room* when the new positions of the guests have all been computed. Using indices *j* and *k,* for example, the routine will examine *room*(*j,k*) and display whatever character is stored there at a corresponding position on the screen.

Moving the *i*th guest is in itself a somewhat complicated operation because his or her new position will depend on the positions of all the other people and the distances the *i*th

© 1987 SCIENTIFIC AMERICAN, INC



Eight stages of the party

© 1987 SCIENTIFIC AMERICAN, INC

	A	B	D	M	P	S	V	W	T
A	0	15	7	2	6	9	4	12	1
B	8	0	6	4	6	3	2	10	1
D	11	4	0	5	12	2	9	6	1
M	6	9	3	0	10	7	13	5	5
P	3	10	5	14	0	11	7	15	5
S	12	2	4	8	5	0	12	4	1
V	7	8	14	10	4	13	0	3	5
W	6	7	13	6	3	8	9	0	5

Table of ideal distances

guest prefers to be from them. Of course, PARTY PLANNER must include a table showing the distance the guest listed by row prefers to be from the guest listed by column [*see illustration above*]. Such an array, called *ideal,* must be stored in the program as a data file or as a set of data statements. The entry *ideal*(*i,j*) records the ideal distance the *i*th guest prefers to be from the *j*th guest. This may not have the same value as *ideal*(*j,i*). For example, Wally may want to be three feet from Penelope, but Penelope wants to be a good 15 feet from him.

Two arrays, *x* and *y,* store the current positions of the eight partygoers. During any given computational cycle the *i*th guest occupies a position whose coordinates are $x(i)$ and $y(i)$. The *i*th guest's level of unhappiness is computed as shown:

```
presum ← sum
sum ← 0
for j ← 1 to 8
    dist2 ← [x(i) − x(j)]² +
            [y(i) − y(j)]²
    dist ← sqrt(dist2)
    sum ← sum + abs[dist − ideal(i,j)]
```

The *for* loop is run not only for the *i*th guest's current position $x(i)$ and $y(i)$ but also for the eight surrounding positions, yielding a total of nine calculations. Every time a new calculation is made, the value of *sum* is stored in the variable *presum* in order to save it for the purpose of comparison. The variable *sum* is then set to zero and the distance *dist* from the *i*th guest to the *j*th is calculated. The absolute value of the difference between the actual distance and the ideal distance (from the *i*th guest's point of view) is then calculated and added to the ongoing total.

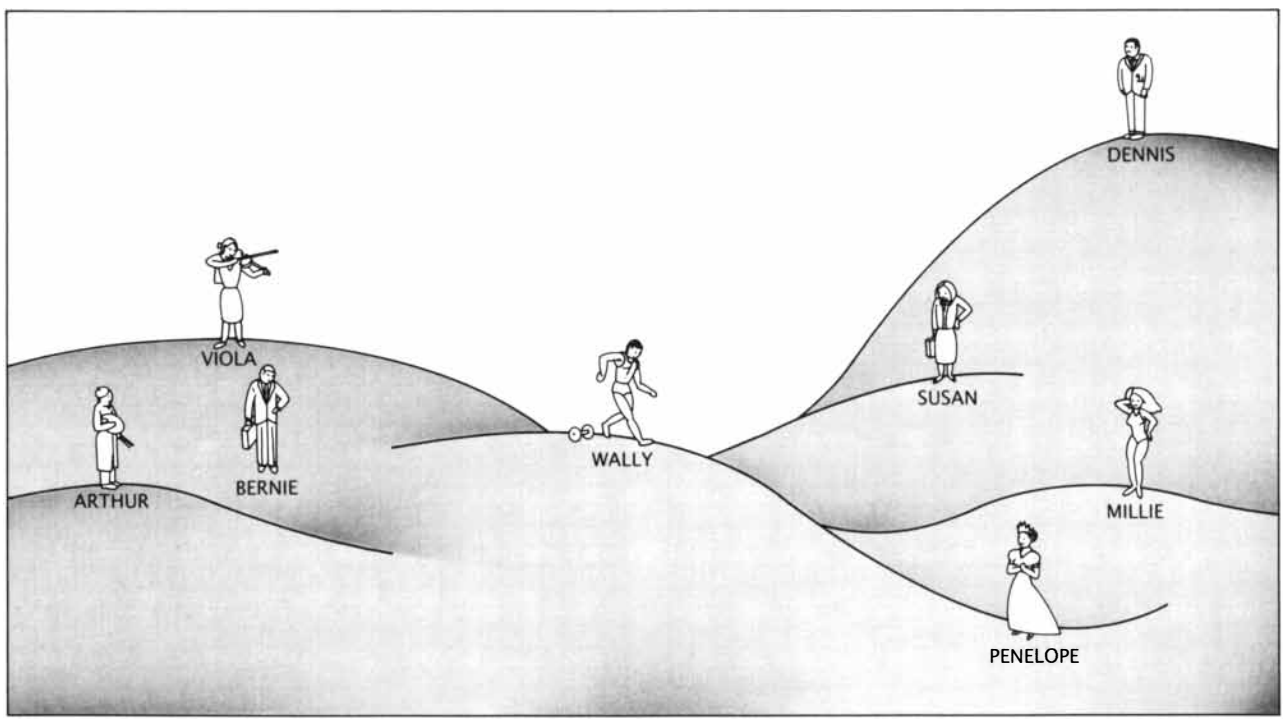
Once a new value of *sum* has been calculated, the program compares it with the previous value stored in *presum.* If the new *sum* is smaller than the previous one, PARTY PLANNER must save the coordinates. The following algorithmic fragment indicates how that might be done:

```
if sum < presum
  then xx ← x(i)
       yy ← y(i)
```

In this way an up-to-date record is maintained of the coordinates that yield the (so far) minimum unhappiness of the *i*th guest. When all nine positions have been calculated, the *i*th guest is moved to the position that yields the least unhappiness. The nine calculations are most simply accomplished by means of a double loop that uses variables *x* and *y* in place of the array elements $x(i)$ and $y(i)$. Thus *x* would go from $x(i) - 1$ to $x(i) + 1$ and *y* would go from $y(i) - 1$ to $y(i) + 1$. A simple way to construct the loop is to add the increments *dx* and *dy* to $x(i)$ and $y(i)$ respectively:

```
for dx ← − 1 to 1
  for dy ← − 1 to 1
    x ← x(i) + dx
    y ← y(i) + dy
```

One further complication should be addressed by party programmers: guests must not be allowed to walk through walls or over the refreshment table. For this reason, every time a new set of coordinate values *x* and *y* is about to be tried PARTY PLANNER must test them against certain bounds that represent the walls and the table. In the array *room* used in the version of the program described here the four walls are at $x = 1$, $x = 30$, $y = 1$ and $y = 20$. If either coordinate of the *i*th guest is equal to one of the corresponding wall coordinates, the calculation of unhappiness must be skipped entirely. Similarly,



Wally the weight lifter descends a hill of unhappiness toward Princess Penelope

© 1987 SCIENTIFIC AMERICAN, INC

the table in my room occupies x coordinates 10 through 15 and y coordinates 9 through 11. If x lies in the first range at the same time that y lies in the second, the guest is in imminent danger of destroying the property of the host. Again the calculation is skipped. Purists might urge that guests not be allowed to walk over each other; this does not occur frequently enough in well-planned parties, however, to be worth the trouble of implementing.

The programmer initiates PARTY PLANNER by selecting starting positions for each guest. The positions can be entered by typing coordinates or by allowing the program to select positions randomly. In the latter case there must be a provision for ensuring that initial positions do not coincide with a wall or the table.

Magi, my cybernetic sidekick, has been experimenting with variations of PARTY PLANNER. Why, he reasons, must everyone have some ideal distance from everyone else? Perhaps some guests will have no feelings at all about certain other guests. When that is the case, Magi inserts a -1 in the appropriate cell of the ideal distance matrix. When the program comes to consider the attraction of that guest for the others, it first tests the value of the ideal distance to the others. If the value is -1, the calculation is skipped; the other guest contributes nothing to the first guest's happiness or unhappiness.

I urge readers to experiment with different room configurations, size of party and array of ideal distances. Gold has even suggested holding what he calls "irrational parties": guests are invited in off the street, as it were. In other words, the ideal-distance array is filled with random integers chosen from an appropriate range of numbers, say 1 through 25.

It is fascinating to experiment with various predetermined guest lists. Sometimes the results are predictable. For example, if all the ideal distances are greater than the room size, the partygoers will all become wallflowers, skirting walls in a vain attempt to avoid everyone else. (I have been to such parties.) If, on the other hand, the ideal distances are small, guests will form a single, tight conversational knot near the refreshments. Hilarious results can be had by setting up an endless cycle of one-way attractions: A loves B but B hates A. At the same time, B loves C but C hates B. The chain continues until it closes in on itself at A again. Depending on how the guests are distributed initially, one may witness an endless chase with occasional romantic eddies and culs-de-sac.

The notion of an ideal distance is not a complete mathematical fiction; behavioral scientists use a similar concept called social distance. In daily interactions with others there is a distance we naturally tend to adopt depending on our relative mobility and our relation to the other person. The study of social distances is called proxemics. Findings include the discovery that social distances are determined in part by culture, role and sex. North Americans, for example, sometimes feel a bit crowded in conversation with people from a country where social distances are smaller. The phenomenon could be studied in a group setting by using PARTY PLANNER. Let half of the guests prefer to be one foot from everyone else and let the other half prefer a distance of three feet. Will they continue to mill about or will some final, stable configuration emerge?

The foregoing question leads us willy-nilly into the realm of dynamical systems. Abstractly considered, a computer party of the kind described here consists of a distance matrix, a two-dimensional cellular space (be it finite or infinite) and a set of tokens that move by turns according to the value of a potential function. From the point of view of a given token (that is, a guest) the space is an undulating plain. The height of any point in this dynamical landscape is the value of the potential function for the token. The value is simply the sum of the differences between the token's matrix distances and actual distances to the other tokens.

The motion of an individual token is analogous to a ball rolling under the influence of gravity; the ball always seeks to move "downhill." By this reasoning Gold sometimes describes the guests at his computer parties as continually seeking valleys of pleasure. At the same time, each token inhabits a different landscape. One guest's valley of pleasure is another's hill of unhappiness [*see bottom illustration on opposite page*].

Given a particular cellular space, whether it is a party room, an infinite plain or a wraparound, toroidal space of the kind so often encountered in the recreations of this department, the questions asked by the abstract investigator are simple: What distance matrices guarantee the ultimate stasis of tokens? What matrices guarantee eventual periodic behavior? The second question makes sense only for infinite spaces, however. In a finite space some configuration must eventually repeat itself. Thereafter the tokens will behave in ruthlessly periodic fashion. In infinite two-dimensional spaces what matrices will never produce periodic behavior?

In a geographic context, Gold has developed a particularly interesting version of PARTY PLANNER. Invite the eight largest cities in the U.S. to a computer party (in a large room). In other words, let the ideal distance between each pair of guests reflect the distance between the corresponding cities. Amazingly, when such guests are given a random initial placement and are allowed to roam freely according to the rules of the game, they usually end up in correct geographic positions in relation to one another! Readers who try this experiment may be momentarily puzzled by the emergence of a map that is reversed or tilted from its usual orientation, but the emergence is undeniable. Presumably the phenomenon occurs for any set of distances between points on a plane.

As an artist, Gold pursues what he calls algorithmic symbolism. PARTY PLANNER emerged not from Gold's need to plan parties but from his need to paint pictures. He writes: "These paintings were to consist of a number of objects such as oranges, apples, violins, hammers, etc. I wanted to paint [them] so that similar objects were close together on the canvas while dissimilar ones were far apart. I had constructed a matrix indicating the ideal distances each object would have from [every] other but discovered it was no easy task to find such an arrangement by eye.... I began to wonder what would happen if I could animate...the objects and let them wander around seeking their own positions. I was not prepared for the beautiful dance they performed as they sought to minimize their own unhappiness. The party planner was born!"

For readers who prefer purchase to perspiration, Gold has diskettes available for the IBM PC and compatible computers as well as a version for the Apple series. His address is 815 North Gardner Street, Los Angeles, Calif. 90046.

The reprise for this column and for all future ones has been moved from three to four months from the month of publication. At that time I shall be delighted to devote a brief social page to the most interesting parties held by readers.

© 1987 SCIENTIFIC AMERICAN, INC

BOOKS

Aerial Armageddon, young Einstein revealed, color's chemistry, high-altitude physiology

by Philip Morrison

THE RISE OF AMERICAN AIR POWER: THE CREATION OF ARMAGEDDON, by Michael S. Sherry. Yale University Press ($29.95).

Just before midnight on March 9, 1945, bombers appeared over unprepared Tokyo. They were no strangers by day, but in blue skies they were remote, gleaming, unreal. Now they droned low, black and menacing; later that night they seemed like silver moths drawn to the towering inferno they themselves had wrought. Although from the ground they seemed to fly at random, there was a plan. First a big X was carved out in flame across the dense residential districts, and then for three or four hours that mark was "fed and broadened" as Gen. Curtis E. LeMay's 300 heavy aircraft poured jellied gasoline and magnesium into the avid flames. This time the spring winds were not divine but infernal; the pillar of fire swept along the surface, consuming 16 urban square miles—home to nearly one and a half million people. Perhaps as many as one in every dozen of them died that night.

Flight was the only recourse, escape in urgent silence, an escape cut off for many by debris and panic, by smoke and fumes, radiant heat and superheated vapors. Some inviting waters were not cool; they boiled many a hapless refugee. Some tenderly bundled babies erupted into flame right on a fleeing mother's back. The grotesque keloid scars of many survivors had hardly formed before the first million citizens abandoned Tokyo.

On March 11 LeMay sent his elated crews against Nagoya; as the nights passed they went on to Osaka, Kobe and back to Nagoya again. By then ground and air crews were exhausted, like the depots of incendiaries at the airfields in the Mariana Islands, and the new "strategic campaign had to go on hold."

It had been a fateful command decision. The bombers had first assembled in October under Gen. Haywood Hansell on the long runways in the Marianas. Efforts to bomb with precision the aircraft-engine plants in Japan had achieved little: the weather was foul; the jet stream, a head wind of unprecedented strength, reached speeds as high as 150 miles per hour; the radar at altitude was hard to read. In early January strategist Hansell was sacked and the 21st Bomber Command was put under Curtis LeMay—tough, cigar-chewing, blunt, the unflinching field commander. In early March, anticipating a probing visit by the analytic deputy chief of staff from Army Air Force headquarters in Washington, LeMay acted.

All at once he decided to bomb by night, to fly at low altitude where there were no hurricane winds, to accommodate more payload by dispensing with defensive guns and gunners (after all, there were few fighters and poor flak). He would send in the full bomber stream, each B-29 heavy with six tons of all-incendiary load. Now there would be wide targets, entire cities to burn under the spring winds.

LeMay's command decision was the center of the intimate articles that ran, one a week, in the *New Yorker* of June, 1945; they were written by the PR officer of the 21st Bomber Command, St. Clair McKelway, who had been an editor of that magazine. He "did not find it contradictory," Michael Sherry now writes, "to say that the 'incendiaries were dropped with precision.'" Fire attack followed on attack. By the end of July, B-san (Mr. B), flying in processions of up to 700 aircraft, had visited and set aflame 60 cities in Japan, large and small, causing deaths at least equaling in number the quarter of a million killed by the two atomic bombers of August. It is even possible that twice that number died.

Bombers cannot be dispatched by a commander's will as easily as were the ready cavalrymen of the Light Brigade, sabers in hand. An industrial engine of war requires elaborate preparation. Years earlier the organic chemists of Harvard had found a cheap way to jell gasoline, the chemical manufacturers had produced the stuff in shiploads and the architects had prepared Japanese houses on the desert proving grounds of Utah. The aircraft designers of prewar years had also been at work. The proposal was in fact no mere option: the documents of the analysts argue persuasively for the scheme. Headquarters had planned the attack carefully; it also preferred that the field commander grasp the weapon and with it the accountability.

Yet LeMay himself believed he had made the critical decision. He even found a clear justification out of his own experience. "I'll never forget Yokohama," he wrote after the war. "That was what impressed me: drill presses. There they were, like a forest of scorched trees and stumps, growing up throughout that residential area." So had it been from the first, a feeder system for the war factories of Japan. By 1944 the Japanese, however, had moved and consolidated production; they "almost eliminated home industry in their war economy," the U.S. Strategic Bombing Survey concluded later. Housing was the "physically the most vulnerable but socially the least effective component of a city." This is hindsight, but even in foresight there was knowledge. What was known pointed to the effectiveness of shutting down transport by sea and land, mining the harbors, sinking the ships, attacking rail and power lines and terminals. There were rationales for sowing fire; none seem weighty enough for the awful events.

It is difficult to deny the author's claim. What LeMay had done "was to bring Americans the pleasure of revenge in the guise of military necessity." It was not the general's decision to set Tokyo aflame that was influential; it was the momentum of his success. By August apocalypse appeared in embryo: two nuclear bombers carried ruin to two cities, cities foreseen and husbanded for years by Gen. "Hap" Arnold, commander of the Army Air Force during World War II, as undamaged test targets. "Which cities should be spared was a problem," he wrote. Now the problem has been generalized.

Michael Sherry is a young historian at Northwestern University. He has

© 1987 SCIENTIFIC AMERICAN, INC

written a brilliant, biting and yet not ungenerous study of a momentous idea, the idea of war from the air, and he has filled in its eclectic origins and its complex sociology, civil and military. Although more than half of the book treats World War II, mainly the air war against Japan, he begins earlier. The first chapter opens with the days of Mark Twain and H. G. Wells, a time that soon saw air war both as terrifying and as attractive because of its presumed swiftness. For Britain the bomber was first seen as a terror; for America, a kind of hope. Daylight precision bombing by unescorted "flying fortresses" was put forward as an American style of war, crisp, even ethical.

It is Sherry's aim to explain how such events happened as a guide to what could come next—to us all. His preface makes it plain that ideas alone cannot suffice; the tangible is required for judging the interplay of idea and action. The striking illustrations are his "visual essay" on American imagination; they are cartoons, posters, advertisements and a few terrible photographs of damage. Yet he does not include many of the tables and estimates that throw light on whether men could indeed do what they imagined. To be sure, he sees the detached style of the analysts of operational catastrophes as in itself a contributor to the fanaticism of technology in war, the exaltation of means before ends, when "satisfaction of organizational and professional drives loomed larger than the overt passions of war."

The 20th century can be seen as an undecided race between the originality and untiring zeal of those who organize war from the skies and the stubborn, inarticulate urge of Everyman to survive. Hare and tortoise?

THE COLLECTED PAPERS OF ALBERT EINSTEIN, VOLUME 1: THE EARLY YEARS, 1879–1902. John Stachel, editor, with six associates. Princeton University Press ($52.50). English translation of the German texts of Volume 1, Anna Beck, translator, and Peter Havas, consultant. Princeton University Press (available only to buyers of the documentary edition, paperbound, $22.50; microfiche, $10; $1.50 for postage and handling in the U.S. and Canada).

A thriving scholarly industry centered in the University of Cambridge has for years documented the lives and works of Isaac Newton and Charles Darwin. Although it is under some linguistic handicap, American competitiveness here asserts itself with a first heavy and handsome volume in condign response, the massive written legacy of our fellow citizen, this century's symbol, at once its hero, victim and unintending culprit.

Einstein once explained that because he had never deferred to authority in his youth, Providence punished him in the end by making him quite an authority himself. This first volume uniquely skirts that punishment by showing us the young Einstein intimately before he had become anything of an authority at all, even before the work of the *Wunderjahr* 1905. It meticulously presents and annotates 134 documents, although all his published papers, even the two of 1901 and 1902, are held over for Volume 2. Here are the letters to and about him as well as from his hand, exams, records of his schoolwork, legal forms he filed and character reports made by the authorities—even his birth certificate, accompanied by photographs of the period as well as copies of sketches and drawings. The book has all the expert apparatus we expect: map, footnotes to identify the people who are mentioned, helpful comment on the more technical material, a chronology, thumbnail biographies of friends, family and teachers and an excerpt from a reminiscence his younger sister wrote in 1924. The texts are in German, although all the editorial apparatus is in English. An accompanying paperback gives a literal translation of the texts but does not repeat any of the editorial matter.

The second item is Einstein's mother's note to her sister, written as he finished second grade at the local public school in Munich. "Yesterday Albert got his grades, once again he was ranked first." The myth of the backward child is plainly false, although his sister had heard that he had been slow in learning to speak; it is certainly true that he was not a top student at all levels and times. He listened to drummers near and far. Young Einstein left Munich at 16, before finishing secondary school, to seek more education in Switzerland and probably to avoid conscription. His family had already moved for business reasons to Italy. The rest of the events in this volume take place in the Swiss setting, except for visits and holidays in Milan, Genoa, Pavia and the mountains.

Einstein failed the general part (but probably not the scientific part) of the entrance examination for engineering at the ETH in Zurich (the Federal Polytechnical School). He had attempted the exam even though he was two years younger than the age of regular admission. He was advised to spend an additional year at the cantonal school in Aarau, a secondary school known for its liberal and secular bent, and good in science.

That year is endearing to read about, and Einstein would recall it fondly all his life. The cantonal inspector reported on an examination of the violin students. Bowing technique seemed a little stiff here and there, but "one student, by name of Einstein, even sparkled...rendering...a Beethoven sonata with deep understanding." Here too Albert fell in love for the first time, with Maria, the young daughter of the family he lived with; his host was a teacher of Greek at the school. "How different it is," he wrote to her while he was away on a visit to his family, "to play a simple, sweet little song with one's sweetheart than to overcome...an admittedly difficult sonata with straight as ramrod, decked-out Pavia ladies." He broke off the innocent romance about a year later. It was in Aarau too that Einstein, not yet 17, wondered what would happen to a light wave if one ran after it fast enough to catch the thing—"the first childish thought-experiment that had to do with special relativity," he would recall at 60.

Nearly a third of the text transcribes two Einstein notebooks that record a year of physics lectures by H. F. Weber, his ETH professor. The student Einstein plainly enjoyed experiments and their results, and he was particularly pleased when a theory would unify empirical results that had seemed to be unrelated. The biographical sensation of the volume is a set of 50 documents never before published, letters exchanged between Einstein and another physics student at the ETH, Mileva Marić from Zagreb. Their letters are funny, tender, intimate and practical, with much comment on physics—if more on jobs and apartments. The lines are full of diminutives; he is Johonzel, little Johnny, for reasons unknown; she is usually Doxerl, a form of Dolly.

Mileva bore to Albert a daughter they called Lieserl, born in Serbia a year before the two were married in Bern, a match much against the Einstein family's wishes. "This Miss Marić is causing me the bitterest hours of my life," the angry mother-in-law wrote. The fate of the baby girl is not known; she may have died in Serbia of a childhood infection. Al-

© 1987 SCIENTIFIC AMERICAN, INC

bert and Mileva had two sons in the marriage, which lasted for nearly 20 years until divorce.

The young physicist tried for every kind of post, "honoring all physicists from the North Sea to the southern tip of Italy with my offer," as he put it. He advertised private lessons in mathematics and physics "given most thoroughly" by Albert Einstein, diplomate. The fees were a help; better were the lifelong friendships he formed with a couple of interesting students. Finally, in June of 1902, through the good offices of a classmate at the ETH, he got a position at the Patent Office: provisional appointment as Technical Expert third class at 3,500 francs to start. "Ehrat thinks...one cannot live on 4,000 fr. with a wife. But we will prove by deeds how fabulously that can be done!" Salary, young wife and son, degree, leisure to work: it would be done well. A few months earlier he had written to Mileva that he was working "very eagerly on an electrodynamics of moving bodies, which promises to become a capital paper." There was truth in that too.

Some 30 volumes lie ahead. The monument being built is a grand one; here is the entrance, lively with the sights and sounds of youthful life.

ARTISTS' PIGMENTS: A HANDBOOK OF THEIR HISTORY AND CHARACTERISTICS, VOLUME 1. Robert L. Feller, editor. Cambridge University Press and National Gallery of Art ($49.50).

The miniature of the delicate young prince in the orange-red robe, like Claude Monet's quieter oil of a wooded fall scene along the Italian Riviera, might appear to be an unlikely and beautiful distraction among these pages. Most of the other images are colorful polarized microphotographs, gray X-ray-diffraction patterns from powders or crisply graphical infrared spectral traces. Actually both works of art belong of right: they are celebrated samples. That Persian robe is brilliant in red lead, an artificial pigment as familiar to Vitruvius as it was to the painters of the Han. The fall foliage is tinted with cadmium yellow, an expensive product of the 19th-century chemists that is still in wide, if sparing, use by artists today.

The volume is a coordinated collection of 10 recent monographs, one on each of as many pigments, compiled as an authoritative reference for artists, conservators and art historians. For the general reader it is agreeably exemplary both of the versatile powers of modern materials analysis and of the breadth and energy of scholarship. The choice of the pigments is more or less opportunistic (these are recent studies from a variety of expert laboratories in the U.S., Canada and Europe); indeed, a second volume will offer 10 more. The chapters share a common systematic organization including a general introduction, a brief history of use, a section addressing the artist's concerns, color range, permanence, grinding and so on; a long chapter on the diagnostic and revealing details of chemical and physical nature and variation, as well as a review of the appropriate analytical techniques. Finally, there is a careful field guide to specific uses of the materials in art.

The pigments here represent the animal and mineral kingdoms; plants play a subtler role. The analytical studies follow the nature of the substance: electron microprobe and X-ray powder-diffraction analyses are fine for heavy metals. A cross section of the paint at one point within a dramatic old canvas, an *Arrest of Christ,* is shown at hundredfold magnification by three techniques: optical microscopy, scanning electron microscopy and X-ray microtrace at two resolutions. Matthias Stomer's yellow underpaint is lead antinomate, a synthetic pigment called Naples yellow, whose use can be traced back to glass of the Middle Kingdom. It has been lost and rediscovered a number of times across three millenniums.

Microspectrophotometry and thin-layer chromatography better suit the intricacies of the organic world. A Persian invention, a clear Indian yellow used liberally in miniatures from the 16th century until the 20th, was made from the urine of cows fed exclusively on mango leaves! Two ounces of the crude pigment were the daily yield of such a cow; an animal so fed appeared chronically ill but could survive on the regimen for a number of years. The government of India seems to have limited production out of respect for the cows in about 1890; cobalt yellows and synthetic organic yellows slowly took the place of the pigment (which is mostly calcium and magnesium salts of an organic acid derived from an active anthraquinone of the dark green leaf). The old yellow pigment, but not its replacements, fluoresces strongly, converting blue-violet light into yellow or yellow-green; photographs of one page rendered in Rajput style show the yellow areas still yellow under the all-purpling ultraviolet lamp.

The redcoats of the British army, Boston's lobsterbacks, were clothed in fabrics given color by the dye derived from the cochineal insect, until it was replaced by azo synthetics in about 1878. The bright crimson of the purest carmine, as the dyestuff is called, is an intensely red powder. A varied set of paintouts are shown here laid in an acrylic vehicle. They are made by treating with iron-free alum the soluble hot-water extract of the insects. The result is an insoluble pigment called a lake. The beautiful colors are fugitive under light, so that nowadays carmine is used mainly for cosmetics; in the U.S. it is the only organic pigment now allowed in eye makeup.

One color photograph, itself a minor work of art, displays the use of simple but apt thin-layer chromatography in the analysis of the carmines. A small sample of each pigment is drawn out across a sintered layer of commercial absorbent polyamide by the diffusion of a few drops of the right fancy solvents. The blots that result are made visible—here as a set of pastel blues and violets with a couple of accenting red tips—by a heavy-metal developer, uranyl acetate. A dozen drifted patterns appear, clearly distinguishing the two carmines, madder, some mixtures and a few control samples as well. Spot tests to be viewed in ultraviolet are also tabulated. This is altogether a valuable reminder of what the expert organic analyst (from the Bayer Analin- und Sodafabrik laboratories) can do today, without electron beam or polarizing microscope. This style comes less easily to the analysts of paintings, but the scheme has identified the carmines in the paintbox of J. M. W. Turner and in paintings by artists from Boucher and Goya to Dufy and de Chirico. Yet it is in faded pre-Columbian textiles that these once dazzling traditional colors are most extensively found in the museums. Here is a splendid example of convergence during the long evolution of a technology that delights the eye.

GOING HIGHER: THE STORY OF MAN AND ALTITUDE, by Charles S. Houston, M.D. Revised edition, 1987. Little, Brown and Company ($19.95; paperbound, $10.95).

Operation Everest II, a simulated 40-day assault on a Himalayan summit, was mounted in a pressure tank late in 1985 at the U.S. Army Research Institute of Environmental Medicine in Natick, a western suburb of Boston. Dr. Charles Houston, a vet-

© 1987 SCIENTIFIC AMERICAN, INC

eran mountaineer, himself no stranger to the real slopes of K-2 and Everest, is a reflective internist with strong physiological interests and a man who demonstrably can both organize and write vividly. He led the team at high pressure (no metaphor!). At low pressure were eight volunteers chosen with care; "all were very fit,...most were climbers." The space was crowded, but not intolerably so: the main decompression chamber was 20 feet by nine, connected to a smaller lock and to a room-size second chamber crowded with instrumentation. The volunteers lacked for nothing except air and room. Theirs was no life in frail tents under the storm winds; they enjoyed snug bunks, showers and the best of food (including the most esteemed ice creams of the day). They did exercise seriously, whether on a treadmill, cycling against a load or on a climbing simulator.

They all first lost appetite, then weight, then sleep, as the air grew slowly thinner. In the air of 20,000 feet all became lethargic and took interest mostly in lying in the sack. They were still able to sustain short bursts of maximum work even at 25,000 feet. They did not stagger or hallucinate even at Everest summit altitudes, as most of the climbers of the real peak remember doing. These indoor mountaineers were not as poorly off as that, yet they were not as well acclimatized toward work in thin air as real climbers are. Was the time too short? It is hard to stretch such an imprisonment out to the several months of real expeditions. Perhaps the stressful work of real climbing is itself a help to acclimatization.

The rich data are still under study: they range from regular muscle biopsy and rates at which injected inert gases are cleared to pressures and oxygen uptake measured by the passage of a small tube all the way from an arm vein right into the heart and lung. The study was successor to a smaller experiment Dr. Houston had conducted in 1946 at the School of Aviation Medicine in Pensacola, Fla.

The expeditions of the pressure chamber are not the chief burden of this compact book, only its newest element. The volume is an up-to-date survey of what is and is not known about the results of reduced oxygen transport and the correspondingly reduced removal of carbon dioxide. The author describes the fundamentals: the fall in oxygen availability step by step from free air to the cell membranes that must admit passively diffusing gas from the tiny capillaries nearby, and at the same time clear waste carbon dioxide. That physiology is explained in simple language. A reader gathers a deep sense of the tangled feedback loops, hormones and enzyme systems that complicate the simplicities of a shortage of the most urgent and essential of all nutrients, oxygen.

In addition to those who dwell there, millions of people visit modest altitudes. (No enduring settlement is known above 17,000 feet, although some Andean mine caretakers have lived for several years at 20,000 feet, and archaeological finds suggest ancient temporary occupation of sites 22,000 feet high.) The clinical record is thick: there are several distinct kinds of altitude illness, and a number of other effects. Cases typical of each are discussed. They come both from the growing literature and from what Houston himself has cannily observed for 50 years.

Acute mountain sickness hits one person in six or eight of those who travel upward within a few days from near sea level to 9,000 feet or more. It is like a bad hangover, with a battery of unpleasant symptoms, more uncomfortable than threatening. It lasts for a day or two. Prevention can be achieved by slowing down or by taking a new drug, Diamox, an inhibitor of the enzyme carbonic anhydrase. According to Houston, the agent sets in motion a combination of actions that begin with an increase in red-cell oxygen affinity; the drug allows more breathing with lower blood alkalinity and with less bicarbonate loss in urine.

Rarer (very roughly one patient among hundreds of visitors to the heights) but much more dangerous is an edema of lungs that have been too rapidly starved of oxygen. Increased arterial pressure stretches the blood vessels and promotes release of substances that allow leakage from fine capillaries. Fluid gathers in the tiny lung spaces, and the process worsens. Several detailed mechanisms are suspected, from platelet clumping to surfactant breakdown. Daily aspirin to reduce the prostaglandin level may be helpful, Houston reports; the best remedy is to get back down. The Indian armed forces anxiously confirmed both the condition and the treatment during the 1962 border conflict in the Himalayas, very soon after the first description of the disease from Peru.

Third is a true brain disorder, marked by insomnia, staggering and hallucinations. Climbers who survived one such attack "said they had seen bulldozers and palm trees on the snow-capped summit." The neat but simplistic explanation is an edema of the brain, which swells because it retains too much fluid. "The brain becomes too big for its box." There is evidence for this, but hardly enough. Compare two photographs of the retina and its blood vessels, at sea level and at altitude. The tiny vessels are doubled in size above 15,000 feet, a universal response to that stress. Go much higher and real hemorrhages appear in the retinal blood net. They seem not to be accompanied by symptoms and usually disappear after a few weeks. Is the eye a window to the mind? Are there such lesions in the brain? Do they matter? The answers are not known.

One can dream of a tentative general model for all of this. The electrolyte ions, sodium and potassium, do not diffuse freely through the cell membrane; they must be actively pumped. Sodium is pumped out of the cell uphill against the concentration gradient; potassium is pushed instead into the cell, where it is normally more abundant. The energy for the pumping demands oxidative synthesis of the cellular fuel ATP. A derangement of pumping rates may be behind most of the watery changes that seem to show up at altitude.

Houston opens with an unusually interesting summary of the long history of our knowledge of breathing, amusingly illustrated. (We miss precise citations for some sources, and there is confusion about early microscopy.) Robert Hooke appears in the story, occupying his usual role as an innovator: in 1671 at a meeting of the Royal Society he decompressed himself in a sealed oaken barrel to the air pressure prevailing at about 3,000 feet. He reported no discomfort except for some pain in the ears. Later the balloonists, audacious if not downright foolhardy, rose without preparation to Himalayan altitudes; sometimes they even survived.

A brief chapter celebrates the remarkable women pioneers of mountain climbing, beginning with Maria Paradis, who climbed Mont Blanc in 1808 in the seventh party to reach that summit. A street in Chamonix bears her name. On the side of physiology there is a suggestion that women are less likely than men to develop the rarer altitude illnesses. Well-trained women, lighter than men but comparably strong, can climb as well as men—or perhaps better.

© 1987 SCIENTIFIC AMERICAN, INC

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

STRATEGIC DEFENSE AND DIRECTED-ENERGY WEAPONS

WEAPONS IN SPACE, VOL. 1: CONCEPTS AND TECHNOLOGIES. *Daedalus,* Vol. 114, No. 2; Spring, 1985. VOL. 2: IMPLICATIONS FOR SECURITY. *Daedalus,* Vol. 114, No. 3; Summer, 1985.

THE STRATEGIC DEFENSE INITIATIVE: PERCEPTION VS REALITY. Wolfgang K. H. Panofsky in *Physics Today,* Vol. 38, No. 6, pages 34–45; June, 1985.

STRATEGIC DEFENSE INITIATIVE: THE POLITICS AND SCIENCE OF WEAPONS IN SPACE. Gerold Yonas in *Physics Today,* Vol. 38, No. 6, pages 24–32; June, 1985.

THE LARGE-SCALE STREAMING OF GALAXIES

THE LARGE-SCALE STRUCTURE OF THE UNIVERSE. P. J. E. Peebles. Princeton University Press, 1980.

EVIDENCE FOR LOCAL ANISOTROPY OF THE HUBBLE FLOW. Marc Davis and P. J. E. Peebles in *Annual Review of Astronomy and Astrophysics,* Vol. 21, pages 109–130; 1983.

A DISTANCE SCALE FROM THE INFRARED MAGNITUDE/H I VELOCITY-WIDTH RELATION, V: DISTANCE MODULI TO 10 GALAXY CLUSTERS, AND POSITIVE DETECTION OF BULK SUPERCLUSTER MOTION TOWARD THE MICROWAVE ANISOTROPY. M. Aaronson, G. Bothun, J. Mould, J. Huchra, R. A. Schommer and M. E. Cornell in *The Astrophysical Journal,* Vol. 302, Part 1, No. 2, pages 536–563; March 15, 1986.

SPECTROSCOPY AND PHOTOMETRY OF ELLIPTICAL GALAXIES: A LARGE-SCALE STREAMING MOTION IN THE LOCAL UNIVERSE. Alan Dressler, S. M. Faber, David Burnstein, Roger L. Davies, Donald Lynden-Bell, R. J. Terlevich and Gary Wegner in *The Astrophysical Journal,* Vol. 313, No. 2, Part 2, pages L37–L42; February 15, 1987.

REVERSE TRANSCRIPTION

FORM AND FUNCTION OF RETROVIRAL PROVIRUSES. Harold E. Varmus in *Science,* Vol. 216, No. 4548, pages 812–820; May 21, 1982.

REPLICATION OF THE GENOME OF A HEPATITIS B-LIKE VIRUS BY REVERSE TRANSCRIPTION OF AN RNA INTERMEDIATE. J. Summers and W. S. Mason in *Cell,* Vol. 29, No. 2, pages 403–415; June, 1982.

TY ELEMENTS TRANSPOSE THROUGH AN RNA INTERMEDIATE. J. D. Boeke, D. J. Garfinkel, C. A. Styles and G. R. Fink in *Cell,* Vol. 40, No. 3, pages 491–500; March, 1985.

THE MOLECULAR BIOLOGY OF HEPATITIS B VIRUSES. Don Ganem and Harold E. Varmus in *Annual Review of Biochemistry,* Vol. 56, pages 651–693; 1987.

ELECTRIDES

OPTICAL SPECTRA OF ALKALI METAL ANION AND "ELECTRIDE" FILMS. James L. Dye, Michael R. Yemen, Michael D. DaGue and Jean-Marie Lehn in *Journal of Chemical Physics,* Vol. 68, No. 4, pages 1665–1670; February 15, 1978.

CESIUM 18-CROWN-6 COMPOUNDS. A CRYSTALLINE CESIDE AND A CRYSTALLINE ELECTRIDE. Ahmed Ellaboudy, James L. Dye and Patrick B. Smith in *Journal of the American Chemical Society,* Vol. 105, No. 21, pages 6490–6491; October 19, 1983.

PHYSICAL AND CHEMICAL PROPERTIES OF ALKALIDES AND ELECTRIDES. J. L. Dye and M. G. DeBacker in *Annual Review of Physical Chemistry,* Vol. 38, pages 271–301; 1987.

MIMICRY IN PLANTS

MIMICRY IN PLANTS AND ANIMALS. Wolfgang Wickler. McGraw-Hill Book Company, 1974.

MIMICRY IN PLANTS. Delbert Wiens in *Evolutionary Biology,* Vol. 11, pages 365–403; 1978.

CROP MIMICRY IN WEEDS. Spencer C. H. Barrett in *Economic Botany,* Vol. 37, No. 3, pages 255–282; July–September, 1983.

MIMICRY AND DECEPTION IN POLLINATION. A. Dafni in *Annual Review of Ecology and Systematics,* Vol. 15, pages 259–278; 1984.

THE SWAHILI CORRIDOR

EARLY SEAFARERS OF THE COMORO ISLANDS: THE DEMBENI PHASE OF THE IXTH–XTH CENTURIES AD. Henry T. Wright in *Azania,* Vol. 19, pages 13–59; 1984.

THE SWAHILI: RECONSTRUCTING THE HISTORY AND LANGUAGE OF AN AFRICAN SOCIETY, 800–1500. Derek Nurse and Thomas Spear. University of Pennsylvania Press, 1985.

ASIATIC COLONISATION OF THE EAST AFRICAN COAST: THE MANDA EVIDENCE. Mark Horton in *Journal of the Royal Asiatic Society of Britain & Ireland,* No. 2 for 1986, pages 202–213; 1986.

HOW CHILDREN LEARN WORDS

THE ACQUISITION OF THE MEANING OF COLOR TERMS: A STUDY OF LEXICAL DEVELOPMENT. Elsa J. Bartlett in *Proceedings of the Stirling Conference on the Psychology of Language,* edited by P. Smith and R. Campbell. Plenum Press, 1977.

DICTIONARIES IN THE MIND. George A. Miller in *Language and Cognitive Processes,* Vol. 1, No. 3, pages 171–185; Spring, 1986.

THE NATURE OF VOCABULARY ACQUISITION. Edited by Margaret G. McKeown and Mary E. Curtis. Lawrence Erlbaum Associates, 1987.

COAL-FIRED POWER PLANTS FOR THE FUTURE

COAL—BRIDGE TO THE FUTURE. Report of the World Coal Study. Carroll Wilson, Project Director. Ballinger Publishing Co., 1980.

COAL GASIFICATION SYSTEMS FOR POWER GENERATION. Seymour Alpert and Michael J. Gluckman in *Annual Review of Energy,* Vol. 11, pages 315–355; 1986.

ENVIRONMENTAL ISSUES AFFECTING COAL TECHNOLOGY: A PERSPECTIVE ON U.S. TRENDS. K. E. Yeager and S. B. Baruch in *Annual Review of Energy,* Vol. 12, pages 471–502; 1987.

THE AMATEUR SCIENTIST

STABILITY OF FLUID INTERFACES OF REVOLUTION BETWEEN EQUAL SOLID CIRCULAR PLATES. R. D. Gillette and D. C. Dyson in *The Chemical Engineering Journal,* Vol. 2, No. 1, pages 44–54; January, 1971.

STABILITY AND OSCILLATIONS OF A SOAP FILM: AN ANALYTIC TREATMENT. Loyal Durand in *American Journal of Physics,* Vol. 49, No. 4, pages 334–343; April, 1981.

COMPUTER RECREATIONS

THE EFFECTS OF PERSONAL SPACE AND TERRITORY ON HUMAN COMMUNICATION. E. T. Hall in *Nonverbal Communication in Human Interaction,* edited by Mark L. Knapp. Holt, Rinehart and Winston, 1978.

MODERN COMPUTER PEOPLE. Activision, Inc., Mountain View, Calif., 1985.

© 1987 SCIENTIFIC AMERICAN, INC

TOYOTA WONDERWAGON!

FAMILY ROOM
Wonderwagon has seating for seven with all the comforts of home. The options are exciting . . . dual sunroofs,** Captain's Chair Package, dual air conditioning, even an ice maker/cooler.

THE WAGON FOR TODAY'S PIONEER FAMILY.

Discover the 1987 Toyota Wonderwagon, designed for active families on the go. More spacious than a station wagon, Wonderwagon drives like a car, turns on a dime. It can carry up to 149 cu. ft. of cargo,* yet it parks in a normal-size garage.

NEW 4-WHEEL DRIVE
The new 4-wheel drive Wonderwagon takes on bad roads and bad weather at the flip of a switch.

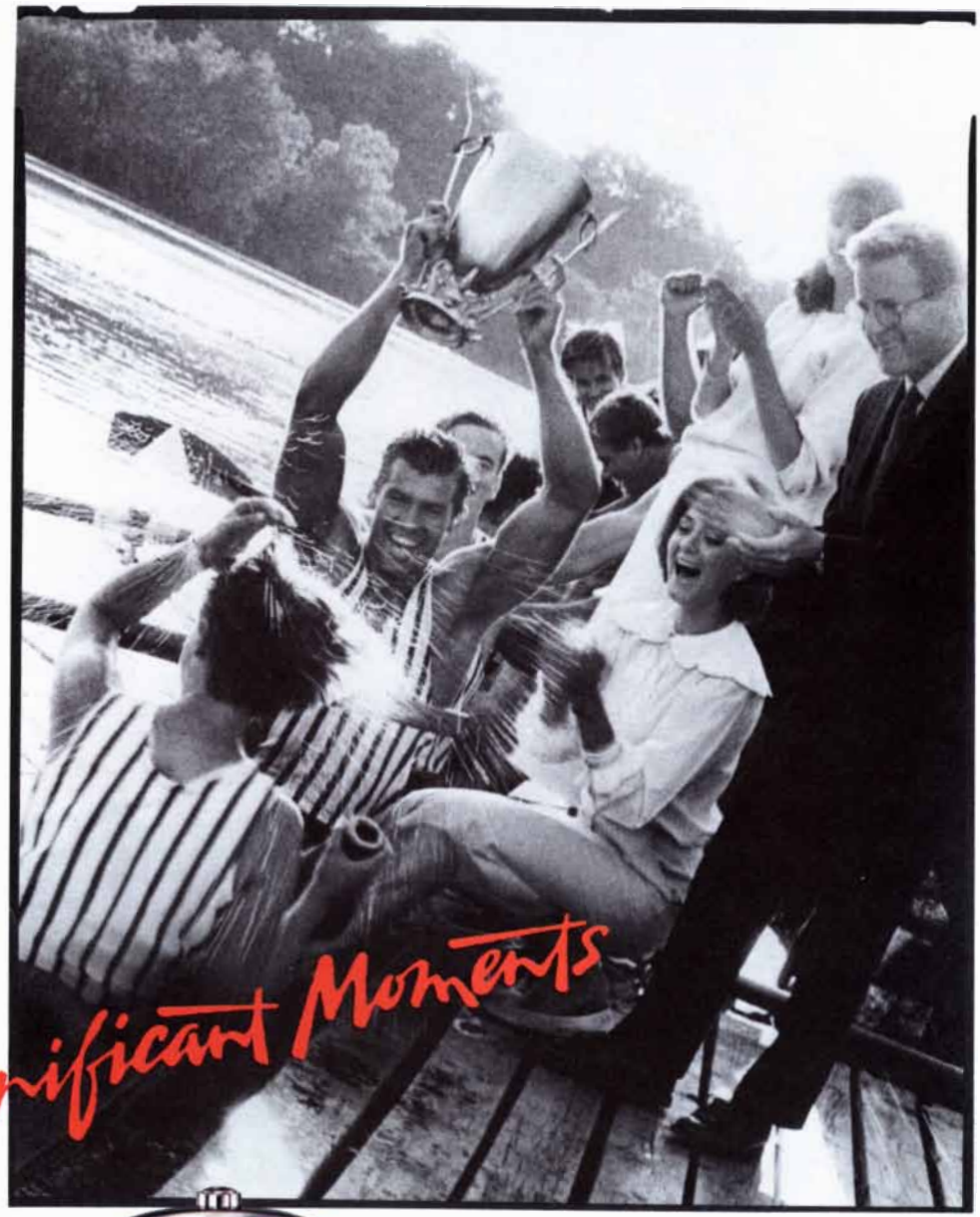
WHO COULD ASK FOR ANYTHING MORE! **TOYOTA**

Get More From Life . . Buckle Up!

*With rear seats removed
**Sunroofs not available on 4WD vans
© 1986 Toyota Motor Sales, U.S.A., Inc.

© 1987 SCIENTIFIC AMERICAN, INC

IT WAS A MOMENT OF RELEASE. FROM YEARS OF PRACTICE. AND A LIFETIME OF DISCIPLINE. OMEGA. FOR ALL YOUR SIGNIFICANT MOMENTS.



Registered Model

OMEGA ALWAYS MARKS SIGNIFICANT MOMENTS.
IN THE OLYMPICS. IN THE SPACE PROGRAM.
IN SIGNIFICANT LIVES LIKE YOURS.
THE OMEGA SEAMASTER. ADVANCED SWISS TECHNOLOGY. WATER RESISTANT TO 30 METERS.
IN STAINLESS STEEL AND 18K GOLD.

Ω OMEGA

MAYOR'S
Fine Jewelers since 1910.

Miami · N. Miami · Coral Gables · Hialeah · Hollywood · Plantation · Ft. Lauderdale · Coral Springs · Pompano
Boca Raton Boynton Beach · Orlando

© 1987 SCIENTIFIC AMERICAN, INC