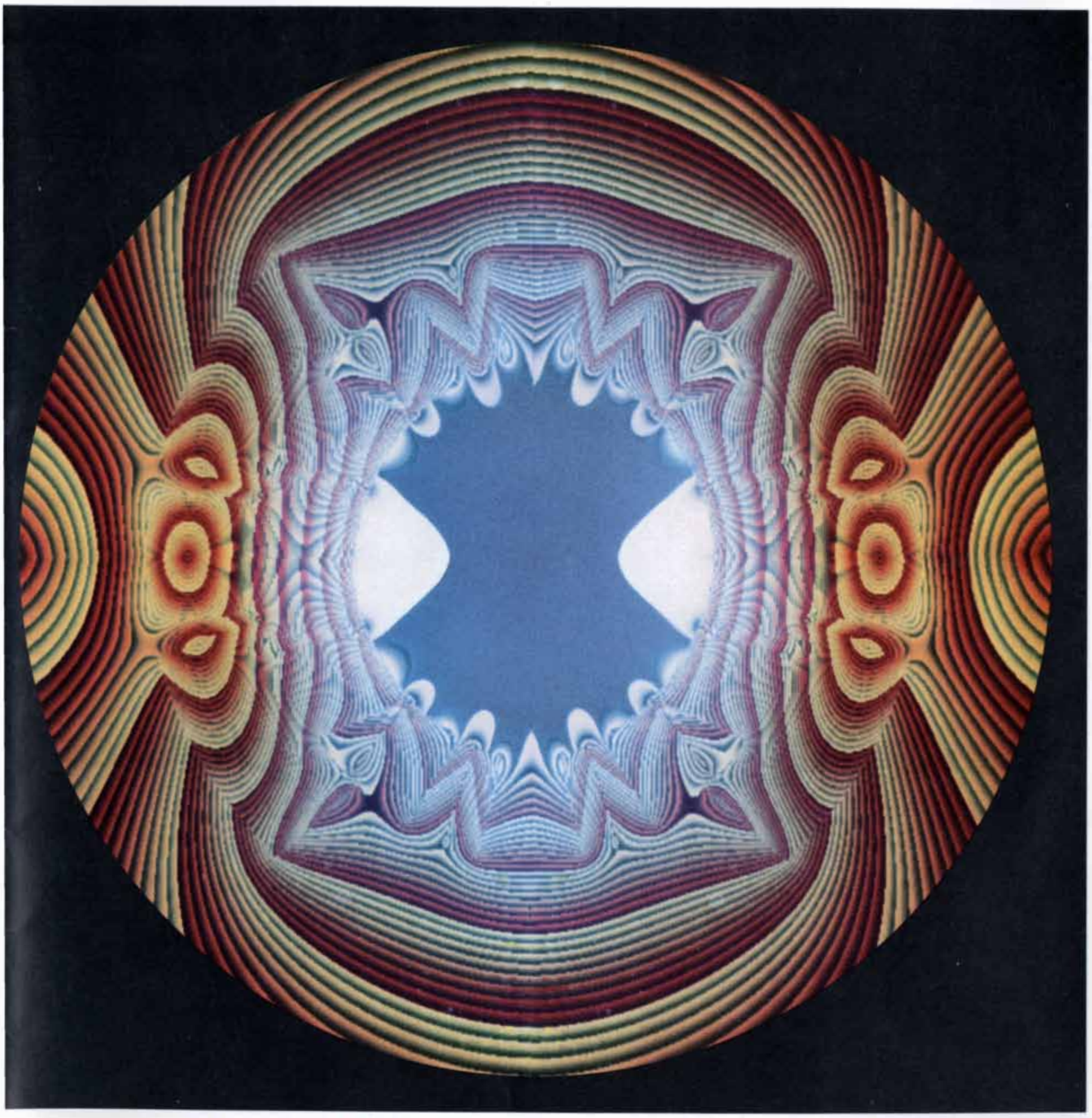
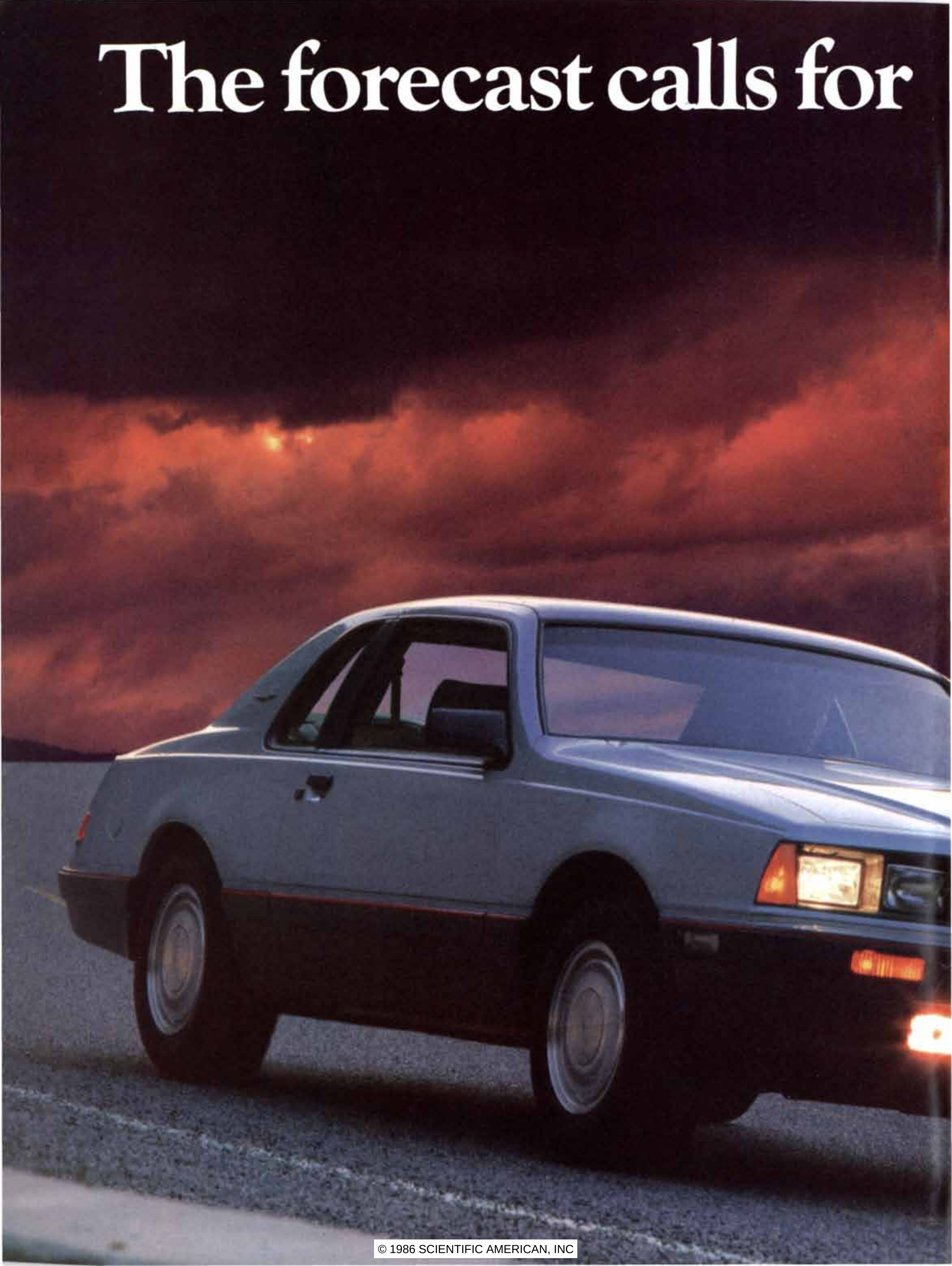
SCIENTIFIC AMERICAN



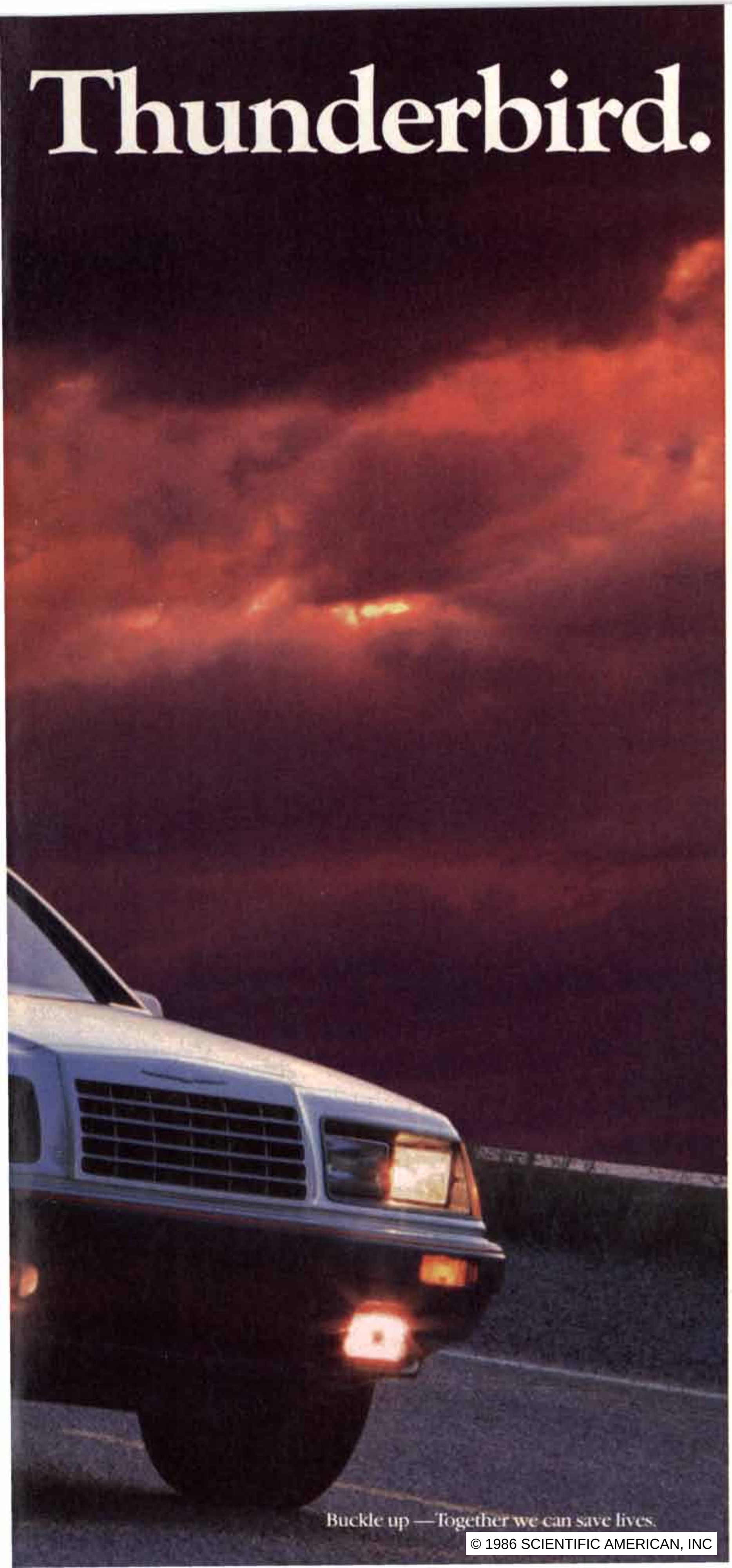
SUPERCONDUCTING SUPERCOLLIDER

$2.50

March 1986


The forecast calls for

Thunderbird.



On the road, an impending storm presents a special challenge—one the driver of a Thunderbird is well-prepared to accept.

Thunderbird's electronically fuel-injected engine provides the power. Steel-belted radial tires and rack and pinion steering provide the grip. And for further stability and road control, Thunderbird's shape helps reduce front and rear lift.

Inside, you'll find the appointments of a true driver's car. Thunderbird's airflow management reduces wind noise and helps keep the windows clean. Seating areas provide lateral support for cornering. And to minimize the time your eyes are off the road, the instrument cluster provides vital information at a glance.

Of course, Thunderbird does have its limits; it can't predict the weather. It can, however, make dealing with a storm a little easier. You can drive a Thunderbird at your nearby Ford Dealer. Have a nice day.

3-Year Unlimited Mileage Warranty.

The new 3-year unlimited mileage warranty covers major powertrain components on 1986 Ford cars. Warranty is limited and certain deductibles apply. Ask to see the 3-year unlimited mileage warranty when you see your Ford Dealer.

Best-built American cars.

"Quality is Job 1." A 1985 survey established that Ford makes the best-built American cars. This is based on an average of problems reported by owners in a six-month period on 1981-1984 models designed and built in the U.S.

Have you driven a Ford... lately?

Ford

maxell

Established 1845 SCIENTIFIC AMERICAN March 1986 Volume 254 Number 3

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1986 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, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

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 del Lauro, 14,
20121 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

400Ptas/copy 4400Ptas/year $33/(abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica S.A.,
Calabria, 235-239
08029 Barcelona, Spain

POUR LA SCIENCE

25FF/copy 245FF/year 290FF/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

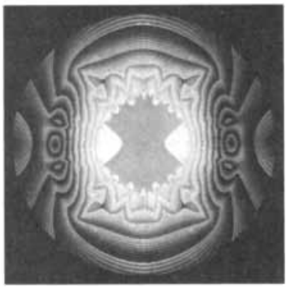
1.40RMB/copy 16RMB/year $24/(abroad)
Editorial, subscription correspondence:
ISTIC-Chongqing Branch, P.O. Box 2104,
Chongqing, People's Republic of China
Advertising correspondence:
C. John Kirby, Advertising Director
SCIENTIFIC AMERICAN
415 Madison Avenue, N.Y., N.Y. 10017

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
Advertising correspondence:
C. John Kirby, Advertising Director
SCIENTIFIC AMERICAN
415 Madison Avenue, N.Y., N.Y. 10017

TUDOMANY

98Ft/copy 1,176Ft/year 2,100Ft/(abroad)
Editorial correspondence:
TUDOMÁNY
H-1536 Budapest, Pf 338
Hungary
Subscription correspondence:
"KULTURA"
H-3891 Budapest, Pf. 149
Hungary
Advertising correspondence:
C. John Kirby, Advertising Director
SCIENTIFIC AMERICAN
415 Madison Avenue, N.Y., N.Y. 10017



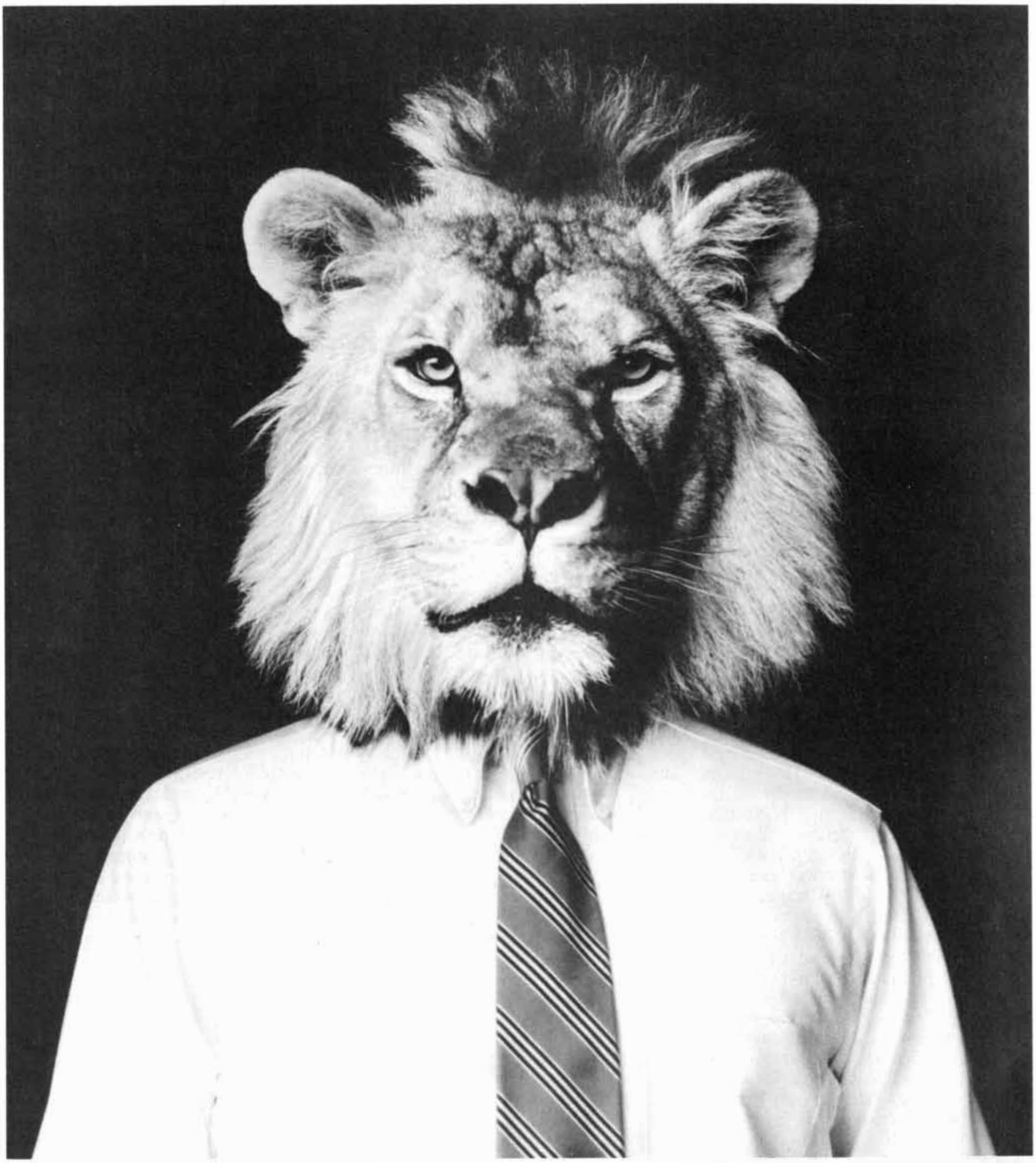
THE COVER

The illustration on the cover is a computer-generated image of the magnetic field that would be generated in the inner core of a superconducting magnet at the proposed Superconducting Supercollider (SSC) (see "The Superconducting Supercollider," by J. David Jackson, Maury Tigner and Stanley Wojcicki, page 66). If construction is approved, the SSC would be the world's largest particle accelerator, and it would enable physicists to investigate matter at previously unattainable scales of length and energy. Superconducting magnets are a key element in the design of the machine. They are to bend two counterrotating beams of protons into a ring some 52 miles in circumference and focus the two beams before they are made to collide. The computer-generated image is color-coded to show the field intensity of one of the bending magnets in cross section, as it might appear to a proton in the evacuated beam pipe. The protons in the pipe pass through the central, blue part of the cross section, about four centimeters in diameter, where the intensity of the field is greatest and most uniform. The intensity of the field in the magnetic coils surrounding the pipe varies rapidly from point to point (*blue, white, purple*). The intensity of the field decreases in the iron yoke surrounding the coils (*orange, brown*), and it is essentially zero outside the core (not shown). The image was made with the computer program POISSON by Shlomo Caspi and Michael Helm of the Lawrence Berkeley Laboratory.

THE ILLUSTRATIONS

Cover image by Shlomo Caspi and Michael Helm

Page	Source	Page	Source
14–21	Edward Bell	89	Stephen Cass
32–39	Ilil Arbel	90–94	Jerome Kuhl
41	Robert H. Eather, Boston College	97	Kenneth A. Wardle, University of Birmingham
42–47	Ian Worpole	98–102	Tom Prentiss
49	Jon C. Lewis, Bowman Gray School of Medicine, Wake Forest University	105	Jon Brenneis
50–54	Alan D. Iselin	106	Hank Iken, Walken Graphics
67	James Kilkelly	107	Peter R. Cavanagh, Pennsylvania State University
68–76	Gabor Kiss	108	Hank Iken, Walken Graphics
79–80	James Kilkelly	109	Jon Brenneis
81	Andrew Christie	110	Hank Iken, Walken Graphics
82–83	Patricia J. Wynne		
84	Andrew Christie	113–118	Michael Goodman



LETTERS

To the Editors:

"The Development of Software for Ballistic-Missile Defense," by Herbert Lin [SCIENTIFIC AMERICAN, December, 1985], is the latest of several articles by authors who predict that the job is likely to be impossible. These critics encourage the image of a monolithic program of 10 million or more lines of code that must operate flawlessly on first use, in spite of inadequate testing and "unknowable" requirements, otherwise the entire defense system will be rendered useless.

Although the Strategic Defense Initiative (SDI) will test our capabilities, there are reasons to be optimistic. Battle management and other functions can be implemented with a number of relatively uncoupled programs that individually are modest in size. This approach relieves cost and complexity (scaling, or the relation of level of effort to program size, will be more nearly arithmetic than exponential), limits the detrimental effects of code modifications and adds to system robustness.

An SDI system cannot be expected to operate flawlessly. The architecture can be compartmentalized, however, employing multiple sensor and weapon systems for each of several defense layers, with minimum linkage between the layers and between systems within a layer. The resulting protection against single point failures can be enhanced by the use of multiple, dissimilar software codes for key functions.

The SDI will not be the first development in which fully stressed operational testing was not possible. The development of strategic offensive weapons and the landing on the moon are other examples. The historically successful process of incremental system validation, in which increasing proportions of the actual system are combined with increasingly realistic testing scenarios and environments, is particularly compatible with the evolutionary development planned for an SDI system. The surveillance mode will operate continuously, and other functions such as the tracking of boosters and the reconfiguring of fictitiously damaged communications networks can be routinely and nonprovocatively exercised. The armed forces and the National Aeronautics and Space Administration regularly use simulations of stressed and unanticipated scenarios in which "Red teams" confront a system with diabolical combinations of contingencies.

Most of the SDI requirements are knowable, having to do with weapons, sensors, communications and other SDI elements that are of our design, not the Soviets'. Although Red teams cannot anticipate all the threats, countermeasures, et cetera, that an enemy could deploy (a problem common to all military systems), the laws of physics and Soviet budgets limit even those unknowns to a finite set.

MALCOLM W. JOHNSTON

Cambridge, Mass.

To the Editors:

In considering the details of the argument in "The Development of Software for Ballistic-Missile Defense" it will be useful to draw distinctions among three activities that prevent the development of the "reliable software" referred to in the subheading: problem definition, algorithm development and computer-program design and coding.

The first of those activities includes the identification of the threat. Difficulties that arise in this area always pose a serious limitation to the effectiveness of defensive measures (remember the Maginot Line?), but they are not a problem of software reliability. Yet a very considerable part of Lin's article, including the account of the mistaken classification of an Exocet missile as "friendly," deals with such difficulties. By this logic a medication (software) must be condemned if it is inappropriately prescribed on the basis of a wrong diagnosis (threat identification).

The second activity, algorithm development, is not specific to software either; algorithms are also required, for example, for generating the elevation tables for field guns and determining the heading correction a pilot must apply while flying in a crosswind. Because algorithm faults usually manifest themselves in systematic deviations from the desired process, they are detected early in the test program and are rare (but not impossible) in operational computer programs. The environment of a ballistic-missile defense may be more tolerant of such rare algorithm faults than most other applications are. The multiple phases during which incoming missiles can be detected and destroyed give rise to fault tolerance, however: each phase can use independently generated and tested algorithms. Another mechanism for fault tolerance can be embedded in the software: current programming techniques encourage the inclusion of assertions that can be used to detect faulty algorithms. If an aiming algorithm is correct, for example, the time-to-go to encounter should decrease steadily (assuming a target that does not maneuver). If an error is detected by the assertion, an alternate algorithm can be invoked.

Finally, we come to what may properly be termed the software-reliability issue: the adherence of the computer program as delivered to the requirements imposed on it by the other activities I have mentioned. The two primary techniques for fault avoidance in this area are analytical verification (proof of correctness) and testing. Their limitations are stated correctly in the article, but some mention should also have been made of the steady improvements in their effectiveness. Moreover, omitting any reference to software fault tolerance results in a significant misstatement of the capabilities for producing reliable software.

H. HECHT

Los Angeles, Calif.

To the Editors:

In "The Development of Software for Ballistic-Missile Defense" Lin cites an incident in the Falkland Islands war in which the British destroyer H.M.S. *Sheffield* was sunk by an Exocet missile. Lin writes that "the ship's radar warning systems were programmed to identify the Exocet missile as 'friendly' because the British arsenal includes the Exocet." It should be noted that the British government has officially denied this. According to the Ministry of Defence, "the computers on board H.M.S. *Sheffield* functioned effectively prior to and during the attack; allegations that the Exocet missile was mistakenly identified as friendly are without foundation."

MARTIN J. MOORE

Crestview, Fla.

To the Editors:

Mr. Johnston does not seem to realize that the SDI is not just another project. If the goal of the SDI were simply to reduce (rather than eliminate) the ballistic-missile threat or to provide a defense for U.S. missile silos, his optimism might have some basis in reality. But the stated goal of the President's SDI program is the wholesale defense of population against the threat of nuclear missile attack, at a level of reliability known in advance to be sufficient to render these weapons impotent.

It is highly unlikely that an SDI-developed ballistic-missile defense (BMD) would provide such a shield if it were actually put to the test. The fundamen-

DODGE ARIES K
5/50 PROTECTION. STANDARD

THE BEST SIX-PASSENGER CAR YOU CAN BUY.

It's true. In fact, for three years running Aries 2-door, 4-door and wagon have been the best six-passenger cars you can buy. That's because Aries offers the highest standard EPA mileage rating, the lowest base sticker prices, the lowest comparably equipped sticker prices, and the longest standard warranty of *any* six-passenger cars—our renowned 5 year/50,000 mile Protection Plan.*

It adds up to this simple, straightforward fact: For 1986, Dodge Aries is still the best six-passenger car you can buy. One reason for this continued success is that we keep improving and refining Aries. And that includes options—like this year's new electronically fuel-injected 2.5 liter engine. Another reason is Aries' long list of standards. Things like front-wheel drive. Electronic fuel-injection. Five-speed transmission. Rack-and-pinion steering. Steel-belted radial tires. And over 30 other features. In short, the kind of hard content that has traditionally made Aries such a great value.

So sure, when we say Aries is the best six-passenger car you can buy, it's a pretty bold claim. But it's also true, or we couldn't say it. And it also sums up the philosophy we've tried to build into Dodge Aries from the beginning. That's probably why we've sold over half a million to families just like yours.

Bottom line. If you're looking for a six-passenger car that can't be beat, nobody can beat your Dodge dealer. Because that's where you can buy or lease* a new Aries K 2-door, 4-door, or wagon.

Dodge
DIVISION OF
CHRYSLER CORPORATION

AN AMERICAN
REVOLUTION

*1986 6-passenger model comparisons: lowest base and comparably equipped sticker prices; highest standard mileage—25 EPA city est. mpg; longest warranty—5 years or 50,000 miles. Limited warranty on powertrain and outer body rust-through, excludes leases, restrictions apply. See copy at dealer.

BUCKLE UP FOR SAFETY.

tal reason is simple. As Johnston notes, an SDI system cannot be expected to operate flawlessly. Whereas other systems with less demanding requirements can tolerate small or infrequent flaws, a "small flaw" in a large-scale defense of population can have enormous consequences. Furthermore, no known or foreseeable technology will enable us to know *prior to actual use* the actual level of performance a BMD system would achieve in action. If the best that can be said of a BMD system that is the linchpin of U.S. security is that it *might* work were it put to an actual test, the threat of nuclear missiles has not been eliminated nor have the weapons been rendered impotent.

Both Johnston and Mr. Hecht point to fault-tolerant software (involving weak coupling between modules, independent development of modules and recovery block programming) as at least partial solutions to the problem of software reliability. They are right: a complex system that makes use of these techniques is much more likely to be reliable than a system that does not. But the use of fault-tolerant techniques cannot ensure that the resulting system will indeed be fault-tolerant. Indeed, the history of complex systems (software and others) provides clear counterexamples. For instance, independently written programs developed from the same specifications do not fail independently.

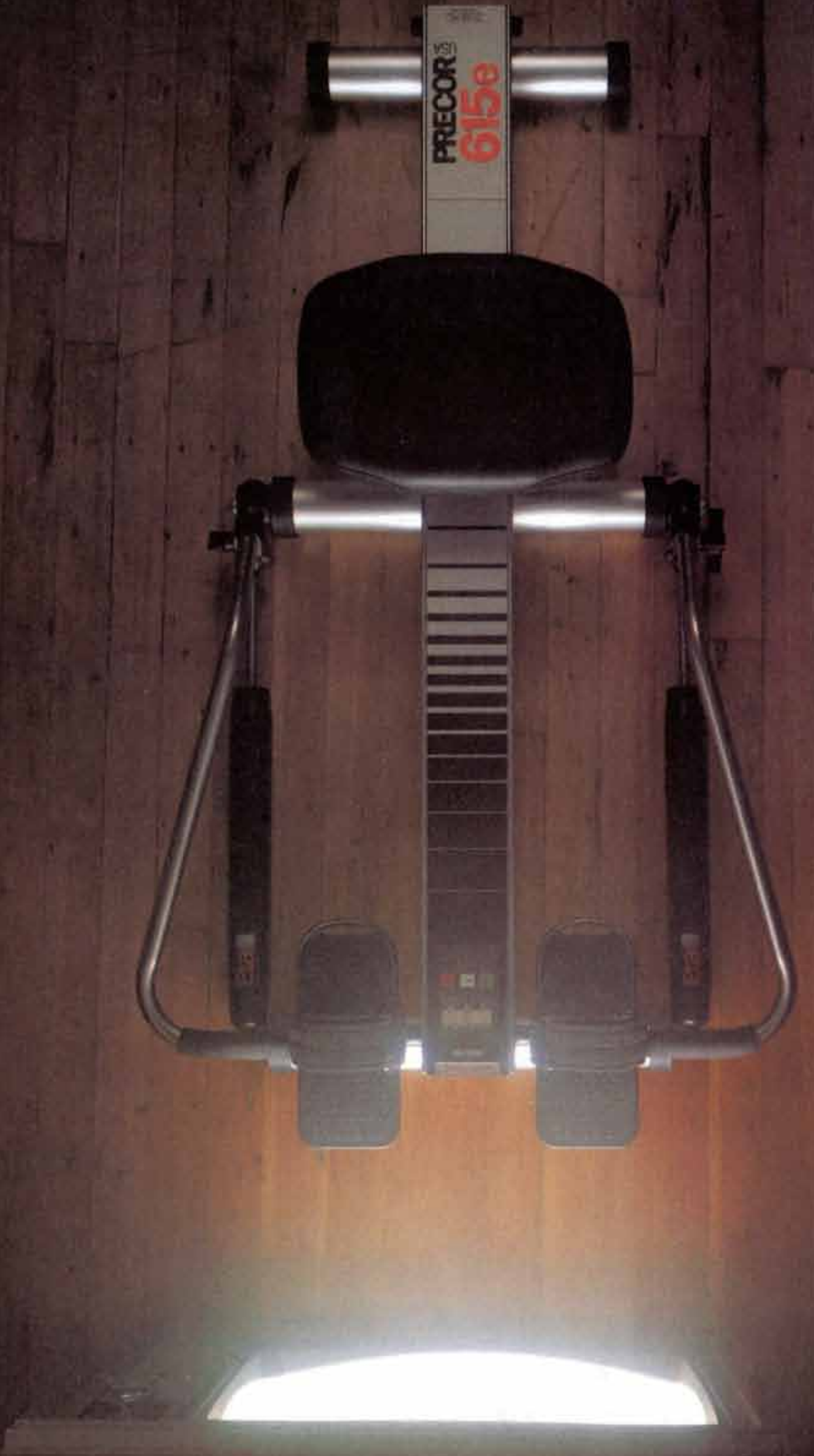
Hecht argues that problem definition and algorithm development are not properly included under the rubric of software reliability. He is wrong: deciding what task is to be done and how it should be done is a very important part of assessing the extent to which computers for BMD will do the job we want them to do. Hecht's assertion does provide a good example of the way very difficult problems are often "solved" by redefining the problem.

Mr. Moore notes my omission of the denial by the British Ministry of Defence of the report about the Exocet/*Sheffield* incident. I had thought that the qualifier "According to one report, the ship's radar warning systems..." would be enough to alert the reader that the statement was not unchallenged, but in retrospect I realize I should have noted the official position as well. In any event, I leave it to the reader to judge the plausibility of the story or the denial.

Finally, a much longer monograph describing the development of software for BMD is available from me for readers interested in further details.

HERBERT LIN

Cambridge, Mass.


PRECOR USA
615e

The Spirit of America

Wyoming Winter by Dick Durrance

Somewhere west of Laramie, men still ride
from dawn 'til dusk. And settle down to a shot of Bourbon
against the chill of the night. Old Grand-Dad still makes that
Bourbon, the only truly American whiskey, just as
we did 100 years ago. It's the spirit of America.

Old Grand-Dad

Kentucky Straight Bourbon Whiskey. 86 Proof. Old Grand-Dad Distillery Co. Frankfort, KY 40601.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MARCH, 1936: "The progress of television in the past few years has been most unsatisfactory from the standpoint of the general public. Now comes a ray of hope. Television has been under the sympathetic eye of the Federal Communications Commission, and in a recent report to Congress the Commission stated that television is practically ready for public use. But—and this is a big 'but'—the Commission went on to state that the various experimenting companies have been working on so many different types of transmission systems that to receive even the experimental transmissions requires a different receiver for each. What is the answer? The Commission states: 'In order to give television service, it is necessary for the different manufacturing companies to standardize their transmissions and produce receivers that can receive all programs transmitted.'"

"There is a great tendency today to add vitamins to foods. The wisdom of this is doubtful, for the ordinary well-balanced diet supplies all the vitamins that human beings usually need. The American Medical Association not long ago denounced the crude and unscientific character of vitamin therapy. It said there was no more reason for people to take varied dosages of several vitamin concentrates incorporated in food or drug products than for them to dose up on any other individual, unrelated dietetic components."

"Travelers by bus or in their own automobiles will soon be able to enjoy custom-made weather wherever they go. Details of a newly perfected system that provides air cooling or year-round conditioning similar to that in use on railroads, in theaters and in stores were outlined recently. The system calls for the use of a refrigerating compressor driven directly by the car motor, which supplies a special refrigerant to an evaporator or a cooling coil."

"Aviation weather service has made notable advances in the development of radio meteorography. The aeronautical meteorologist needs daily and even hourly cross sections of the atmosphere with regard to its pressure, temperature, humidity, wind direction and velocity up to great heights. The ideal method of observation is the sounding balloon equipped for radio transmission, so that there is no limitation as to height and so that information is instantly transmitted to the weather station; in other words, a sounding balloon provided with a 'radio meteorograph.'"

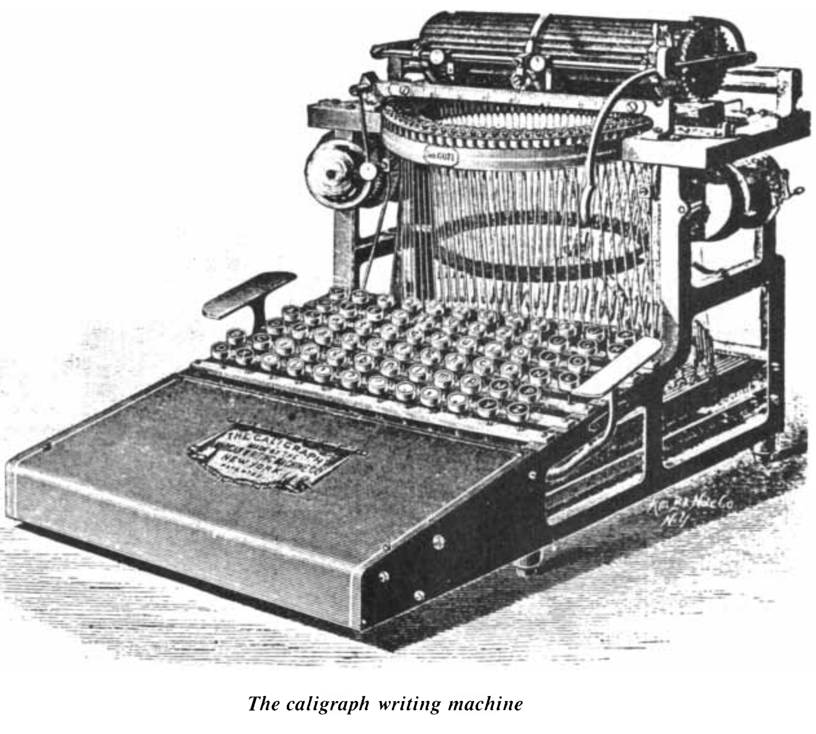
SCIENTIFIC AMERICAN

MARCH, 1886: "Professor Huxley, in his presidential address before the Royal Society, said that 'of late years it has struck me that those who have toiled for the advancement of science are in a fair way of being overwhelmed by the realization of their wishes. It has become impossible for any man to keep pace with the progress of the whole of any important branch of science. It looks as if the man of science of the future were condemned to diminish into a narrower and narrower specialist as time goes on. It appears to me that the only defense against this tendency lies in the organization and extension of scientific education in such a manner as to secure breadth of culture without superficiality; and, on the other hand, depth and precision of knowledge without narrowness.'"

"The growth of the telephone is one of the most remarkable in the history of inventions. In August, 1877, the number of instruments in use in this country was only 780, while in February, 1885, there were 325,574. The number of exchanges has grown from 100 in 1880 to 782 in 1885. In January last there were 137,223 miles of telephone wire in this country."

"A trial has just been made at Portsmouth of an installation of the electric light, which has been fitted on board the *Imperieuse* by Messrs. Siemens Brothers & Company, who are also about to provide similar installations on board the *Warspite, Edinburgh, Collingwood* and *Rodney.* The lights on board the *Imperieuse* comprise 375 incandescent lamps of 20 candle power, which are disposed so as to illuminate all parts of the ship, and also a couple of search arc lights, placed at the bows and at the stern, which are each equal to the power of 25,000 candles. The initial trial was very satisfactory."

"Ten years ago writing machines were little used, were practically unknown to the great majority of writers and were held by the few who knew something of them to be mechanical toys rather than the great time and labor savers they have since proved to be. Up to 1881, when the American Writing Machine Company of Hartford, Conn., introduced the caligraph, double-case writing machines were incomplete, being so constructed as to compel the operator to shift the carriage by a gratuitous stroke for capital letters and figures. The caligraph, however, prints each character in both capitals and small letters at a single finger stroke."



The caligraph writing machine

THE AUTHORS

RICHARD K. LESTER ("Rethinking Nuclear Power") is associate professor of nuclear engineering at the Massachusetts Institute of Technology. Born in England, he was graduated from the Imperial College of Science and Technology in London. A Kennedy Memorial Trust Scholarship enabled him to come to the U.S. in 1974 for graduate studies in nuclear engineering at M.I.T. In 1977 he took a leave from M.I.T. to spend a year and a half as a visiting research fellow in the Division of International Relations of the Rockefeller Foundation. He got his Ph.D. from M.I.T. in 1979 and joined the faculty there as assistant professor; in 1982 he became associate professor. Lester has written extensively on issues relating to world nuclear-energy development.

EDWARD W. HONES, JR. ("The Earth's Magnetotail"), is a geophysicist at the Los Alamos National Laboratory. His undergraduate and graduate degrees are from Duke University: a B.S. (1943) in mechanical engineering and an M.S. (1949) and a Ph.D. (1952) in physics. He did space-physics research at the Convair Corporation, the Institute for Defense Analyses and the University of Iowa before joining the staff at Los Alamos in 1965. In addition to his work in space physics Hones pursues interests in cosmology and laboratory plasma physics.

RICHARD M. LAWN and GORDON A. VEHAR ("The Molecular Genetics of Hemophilia") are senior scientists at Genentech, Inc. Lawn was graduated from Harvard College in 1969 with a degree in astronomy. After beginning graduate work he switched fields and earned a doctorate in molecular biology from the University of Colorado at Boulder in 1977. Lawn did postdoctoral research at the California Institute of Technology, where he took part in the construction of the first human genomic library, and joined Genentech in 1980. Vehar has a B.A. (1970) from Bowling Green State University and a Ph.D. (1976) from the University of Cincinnati. He began his work on the antihemophilic factor while he was a postdoctoral fellow in the department of biochemistry at the University of Washington. Since 1980 he has been at Genentech.

J. DAVID JACKSON, MAURY TIGNER and STANLEY WOJCICKI ("The Superconducting Supercollider") are members of the Central Design Group for the Superconducting Supercollider (SSC). Jackson, a deputy director of the Central Design Group, is professor of physics at the University of California at Berkeley. He holds a B.Sc. (1946) in physics and mathematics from the University of Western Ontario and a Ph.D. (1949) in physics from the Massachusetts Institute of Technology. He taught at McGill University and at the University of Illinois at Urbana-Champaign and moved to Berkeley in 1967. Tigner, the director of the Central Design Group, is professor of physics at Cornell University. He got a B.S. at the Rensselaer Polytechnic Institute in 1958 and a Ph.D. in experimental physics and electrical engineering from Cornell University in 1962. He has long been involved in the design, construction and operation of particle accelerators in both the U.S. and Europe. Wojcicki is chairman of the physics department at Stanford University. A native of Poland, he received his bachelor's degree at Harvard College in 1957 and his doctorate from Berkeley in 1961. After doing research abroad and at Berkeley he became a member of the faculty at Stanford in 1966. Wojcicki is a deputy director of the Central Design Group.

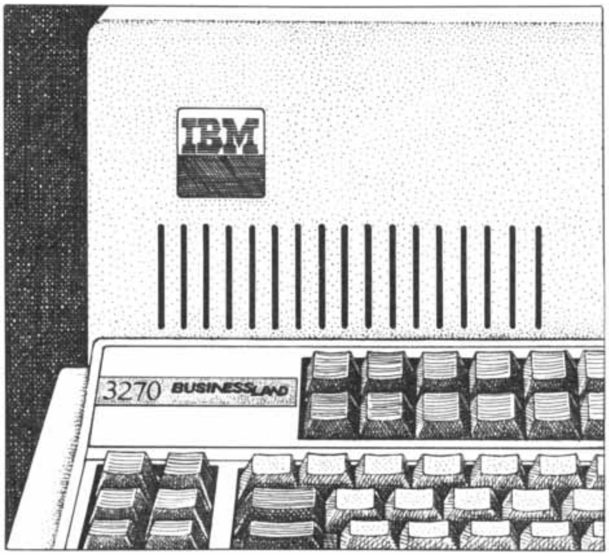
KARL J. NIKLAS ("Computer-simulated Plant Evolution") is associate professor of biology at Cornell University, with appointments in the sections of ecology and systematics and of plant biology. He received his B.S. in biology and mathematics at the City College of the City University of New York; his M.S. and his Ph.D., which he got in 1974, are from the University of Illinois at Urbana-Champaign. He attended the University of London as a Fulbright-Hayes Fellow and became a curator at the New York Botanical Garden in 1975. In 1978 Niklas joined the faculty at Cornell.

RONALD A. FINKE ("Mental Imagery and the Visual System") is assistant professor of psychology at the State University of New York at Stony Brook. He holds bachelor's degrees in physics (1972) and psychology (1974), both from the University of Texas at Austin. He earned his doctorate in psychology at the Massachusetts Institute of Technology in 1979. In 1981, after doing postdoctoral work at Cornell University and at Stanford University, he accepted a position as assistant professor of psychology at the University of California at Davis. In 1983 Finke moved to Stony Brook, where in addition to studying mental imagery he does research on distortions in visual memory and the psychology of belief in paranormal phenomena.

GLYNIS JONES, KENNETH WARDLE, PAUL HALSTEAD and DIANA WARDLE ("Crop Storage at Assiros") are members of the team that is excavating and studying the Bronze Age site at Assiros Toumba. Jones, who specializes in the study of plant remains, is a lecturer in the department of archaeology and prehistory at the University of Sheffield in England. She studied zoology at the University of Wales, where she got a B.Sc. in 1971. After working at an archaeological-science laboratory in Athens, she pursued graduate studies in archaeology at the University of Cambridge, where she got her master's degree in 1978 and a doctorate in 1983. Kenneth Wardle, the director of the excavations, is a lecturer in the department of ancient history and archaeology at the University of Birmingham. He studied at the University of Cambridge and at the University of London, which granted him a Ph.D. in archaeology. Halstead, who specializes in the study of animal remains, is a lecturer at Sheffield. Both his undergraduate and graduate studies were done at Cambridge, where he received a B.A. in 1973 and a Ph.D. in 1984. Diana Wardle, who studies the cultural material uncovered at the site, qualified as an archaeological conservator and draftsman at the British Ministry of Public Building and Works, now known as the Historic Monuments Commission. She is director of the Mycenae Project, one of whose functions is to introduce schoolchildren to Greek legend and prehistory. The authors would like to express their gratitude to the British School at Athens and the Greek Archaeological Service for their support.

CHESTER R. KYLE ("Athletic Clothing") is a sports-equipment consultant and adjunct professor of mechanical engineering at California State University at Long Beach. He got a bachelor's degree in mechanical engineering at the University of Arizona and M.S. and Ph.D. degrees in engineering from the University of California at Los Angeles. He was head of the group that designed the bicycles used by the U.S. cycling team in the 1984 Olympic Games. Kyle is cofounder and past president of the International Human Powered Vehicle Association and has designed human-powered vehicles that have set a number of world records.

OUR 3270 WORKSTATION HASN'T FORGOTTEN HOW TO BE A PC.



Unlike IBM®'s 3270-PC, the new Businessland 3270 workstation requires less memory to run.

140K less to be exact.

Which means your personal computer is free to remain a fully functional, fully compatible desktop tool. Even while it's hosting conversations with your IBM or IBM compatible mainframe.

The Businessland 3270 also provides additional functions you won't find on any other 3270-PC. Like 3287 host printer emulation.

It'll even run a variety of emulation software. From multiple session 3270-PC to IRMA™ to IBM 3278/79.

Plus, Businessland 3270 workstations support many popular file transfer programs. Like IBM, IRMA, or any number of IRMA compatibles.

Or if you prefer, you can use the Businessland file transfer program, which we'll include at no extra charge.

And while we're on the subject of expense, Businessland 3270 workstations start at less than $5,000.*

Of course, there are a few things at Businessland we simply cannot put a price on. Like the marketing representatives, technicians, and system engineers who've been specially trained to help you get the most out of your 3270 workstation.

So call or stop by any one of our 66 nationwide locations and let us show you a videotape of the Businessland 3270 workstation in action.

It could be the most unforgettable 3270 demonstration you'll ever see.

BUSINESSLAND®

Where business people are going to buy computers.

Call (800) 228-7463 for 66 Businessland centers nationwide.

*IBM PC XT, one 360 KB floppy, IBM mono display and adapter, 640 KB of memory. Upgrade kits available for your existing IBM PC, XT, AT, and compatibles starting at $1,195. Volume purchase discounts available. IBM is a registered trademark of International Business Machines Corp. IRMA is a trademark of Digital Communications Associates, Inc.

COMPUTER RECREATIONS

How a pair of dull-witted programs can look like geniuses on I.Q. tests

by A. K. Dewdney

There is an old vaudeville comedy routine that pokes good fun at the strong-man act familiar from the circus and the state fair. A heavily muscled man takes the stage with his not so heavily muscled female assistant. The man strains mightily against an enormous weight, and after tremendous effort he manages to lift it above his head. The spectators cheer, but the cheers turn to laughter when the assistant casually picks up the weight in one hand and carries it offstage.

There are two computer programs that leave one with a similar sense of comic deflation over the mental "muscle" allegedly displayed by a high score on the traditional I.Q. test. Both programs perform at or near genius level on two tasks widely used in the tests, the completion of numerical sequences and the perception of visual analogies. Yet both programs are simple to understand, and it is startling to realize just how dull-witted they are.

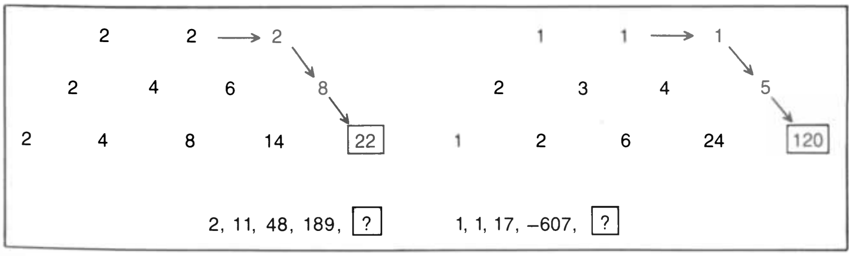
Although I have no wish to offend readers who suppose themselves plentifully endowed with mind stuff, I am twitting the I.Q. test with a serious purpose. The stated intent of the test is to measure intelligence, and few human qualities evoke such pride in their presence or anxiety over their absence as intelligence does. Nevertheless, the concept of intelligence presupposed by the traditional I.Q. test is seriously misguided. The reasoning behind this assessment is cogently set forth by Stephen Jay Gould of Harvard University in his book *The Mismeasure of Man.* What it comes to is this: The traditional I.Q. test rests on the unstated and erroneous assumption that intelligence, like strength, is a single quality of human physiology that can be measured by a graded series of tasks.

Numerical-sequence completion is a good example: What is the next number in the sequence 2, 4, 6, 8, ...? In the sequence 2, 4, 8, 14, ...? In the sequence 1, 2, 6, 24, ...? The percentage of correct responses to a set of such questions measures your "general intelligence," just as a strain gauge measures the weight you can lift and therefore the strength of your arm muscles. Note that if the results of the I.Q. test are to be interpreted as a measure of "general intelligence," there must be some core ability, or some small set of core abilities, that provides an index of what one means by general intelligence. Because the very idea of general intelligence presupposes a strong correlation among the core abilities, the precise kind of graded task adopted by the I.Q. test is relatively unimportant. One task is as good as another.

One of the I.Q. programs presented here is derived from a more elaborate program written by Marcel Feenstra, a student living in Rotterdam. Feenstra's program is called HI Q, and it solves two kinds of numerical problems that often appear on standard I.Q. tests: sequence completion and numerical analogies. Feenstra recently tested HI Q on a number of sample I.Q. tests that appear in a book by Hans J. Eysenck of the University of London, *Know Your Own I.Q.* The I.Q. of HI Q is apparently in the neighborhood of 160. Although the experiment was not exactly a carefully controlled one, it leaves little doubt that the program would score quite well under real test conditions.

The program I have in mind is called SE Q, and it duplicates HI Q's performances on numerical-sequence completion. Readers who write and run SE Q may consider their own numerical intelligence amplified, as it were, by proxy. The main idea of the program is straightforward. When one is given a sequence of numbers and told to find the next number in the sequence, one does not search for the number directly. Instead one searches for the rule that led to the numbers already present. There is a mathematical aside to be made here: For any finite sequence of numbers, there are infinitely many rules that give rise to it. The search thus boils down to finding a simple rule for generating the sequence.

There are just two kinds of rules considered by SE Q: additive and multiplicative. For example, to find the next number in the sequence 2, 4, 8, 14, ..., one might look for an additive rule, and the best way to find the rule is to construct what I call a difference pyramid [*see illustration on this page*]. At the bottom of the pyramid is the given sequence, and the pyramid is built up from bottom to top by finding the differences between successive numbers in the preceding level or row of numbers. Thus the first number in the second row of the pyramid is obtained from the first two numbers in the first row, namely 2 and 4. Their difference is 2, and so 2 is the first number in the second row. Similarly, the other numbers in the second row are 8 − 4, or 4, and 14 − 8, or 6; the second row is the sequence 2, 4, 6. Continuing the same process to a third row of the pyramid gives a sequence with only two numbers, and they are both 2's.



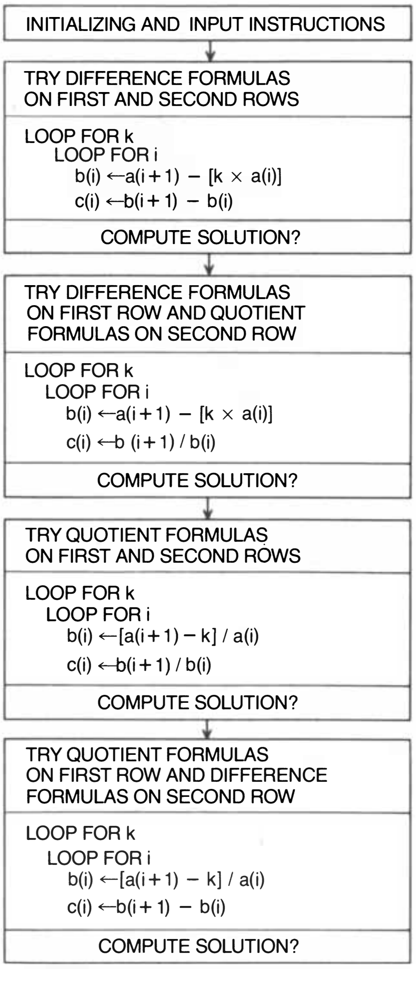
Numerical-sequence completion by the pyramid method. Can the reader solve the lower two?

The equality of all the numbers in some row of the pyramid is the signal, so to speak, to stop building the pyramid upward and to start building it sideways. For example, suppose the third number in the third row is also 2. It is then reasonable to suppose the next number in the second row is obtained from the preceding number, namely 6, by adding 2: the sum is 8. The newly derived number in the second sequence can then be added to the last number given in the first sequence: 14 plus 8 is 22, and 22 is indeed given a perfect score by the test makers. New numbers in each sequence percolate

down the pyramid once a constant sequence is derived at the top.

A great many questions on I.Q. tests about numerical sequences yield to this simple procedure. Readers who have more than a nodding acquaintance with algebra will recognize the signature of a polynomial in the exercise. Any polynomial evaluated for consecutive integers yields a sequence that generates a difference pyramid. Given enough values of the polynomial, a row of identical numbers will eventually top off the difference pyramid. The number of rows needed to build the pyramid up to a constant row, minus 1, is the degree of the polynomial. The sequence 2, 4, 8, 14, which gives rise to a constant row of 2's in the third level of the difference pyramid, is generated by successive values of the quadratic, or second-degree, polynomial $x^2 - x + 2$.

Unfortunately one cannot solve all sequence questions by making difference pyramids. For example, the sequence 1, 2, 6, 24, ... yields a difference pyramid with the numbers 3 and 14 in its top row. The rapid growth of the numbers, however, strongly suggests a geometric series: the consecutive terms of a geometric series are related by multiplication instead of addition. Hence it seems reasonable to construct a set of quotients from the sequence instead of a set of differences [*see illustration on page 14*]. By taking quotients of successive pairs in the sequence 1, 2; 6, 24, ... one obtains the second row in a pyramid, the sequence 2, 3, 4, ... The second sequence hints at an abrupt rule change: the third row in the pyramid must be obtained by taking differences, not quotients. Who can doubt that the intended solution requires a 5 at the end of the second row? The solution itself is thus 120: the product of 24, the last given number in the first row, and 5.



```
INITIALIZING AND INPUT INSTRUCTIONS

TRY DIFFERENCE FORMULAS
ON FIRST AND SECOND ROWS
LOOP FOR k
    LOOP FOR i
        b(i) ←a(i+1) – [k × a(i)]
        c(i) ←b(i+1) – b(i)
COMPUTE SOLUTION?

TRY DIFFERENCE FORMULAS
ON FIRST ROW AND QUOTIENT
FORMULAS ON SECOND ROW
LOOP FOR k
    LOOP FOR i
        b(i) ←a(i+1) – [k × a(i)]
        c(i) ←b (i+1) / b(i)
COMPUTE SOLUTION?

TRY QUOTIENT FORMULAS
ON FIRST AND SECOND ROWS
LOOP FOR k
    LOOP FOR i
        b(i) ←[a(i+1) – k] / a(i)
        c(i) ←b(i+1) / b(i)
COMPUTE SOLUTION?

TRY QUOTIENT FORMULAS
ON FIRST ROW AND DIFFERENCE
FORMULAS ON SECOND ROW
LOOP FOR k
    LOOP FOR i
        b(i) ←[a(i+1) – k] / a(i)
        c(i) ←b(i+1) – b(i)
COMPUTE SOLUTION?
```

Conceptual flow chart for SE Q

The sequence-solving program SE Q attempts to build pyramids by considering both the consecutive differences and the consecutive quotients of successive pairs of numbers in a given row. Even more, it examines successive pairs of numbers in a sequence for more general additive and multiplicative rules. In the additive rule the first member of each pair may be multiplied by a constant k before the usual addition is done, and in the multiplicative rule the constant k may be added just after the usual multiplication. Here is an easy piece for programming novices [*see illustration at left*].

The few simple formulas that give the general rules make up the core of SE Q. Suppose the given sequence has already been assigned to the four variables $a(1)$, $a(2)$, $a(3)$, $a(4)$. To obtain the second row, $b(1)$, $b(2)$, $b(3)$, SE Q tries substituting either a generalized difference, of the form $b(1) \leftarrow a(2) - k \times a(1)$, or a generalized quotient, of the form $b(1) \leftarrow [a(2) - k]/a(1)$. In both examples k stands for any integer in some predetermined range. The program also tries analogous substitutions for $b(2)$ and $b(3)$, each for the same value of k: for $b(2)$ it tries $a(3) - k \times a(2)$ or $[a(3) - k]/a(2)$, and for $b(3)$ it tries $a(4) - k \times a(3)$ or $[a(4) - k]/a(3)$.

The third row, $c(1)$, $c(2)$, is developed even more simply: SE Q tries substituting only simple differences, $c(1) \leftarrow b(2) - b(1)$ and $c(2) \leftarrow b(3) - b(2)$, or simple quotients, $c(1) \leftarrow b(2)/b(1)$ and $c(2) \leftarrow b(3)/b(2)$. Apparently it is rare for sequence-completion questions on I.Q. tests to get more complex than the formulas allow for.

When SE Q develops a pyramid, it tries each generalized substitution for the set of b's with each simple substitution for the set of c's. Conceptually, therefore, SE Q is made up of four major segments. Each segment is a loop with one combination of substitution formulas in it. For example, one such segment of the program first applies the three generalized quotient formulas, of the form $b(1) \leftarrow [a(2) - k]/a(1)$, to compute the values of $b(1)$, $b(2)$ and $b(3)$ that make up the second row of the pyramid. The program segment then applies the two simple difference formulas, of the form $c(1) \leftarrow b(2) - b(1)$, to obtain the values of $c(1)$ and $c(2)$ that make up the third row. The complete set of five formulas in this segment is embedded in what might be called a try-everything loop, in which different values of k are tested. Feenstra recommends allowing k to take on all integer values from -5 to 5.

Within the loop, each time new values for $c(1)$ and $c(2)$ are computed they are tested for equality. If they prove to be equal, their common value is stored in a variable called c and the current value of k is saved in a variable called kk. Just after the loop there are instructions that construct the solution to the original sequence (if one has been found) from the values of c and kk. In the example I am describing one obtains $b(4)$, the new member of the second row, by adding c to $b(3)$. The solution, $a(5)$, is then obtained by multiplying $a(4)$ and $b(4)$ and adding kk to the product.

Two instructions at the end of each loop thus suffice to recover a solution from a successful search within the loop. The instructions appropriate for each loop depend on the formulas used in it, and I shall leave it to those who write SE Q to discover the instructions for themselves. Use a bit of algebra to isolate the variable of interest. In each case, if one of the loops in the program finds a solution, there must be an instruction to print it. The program may then skip the remaining loops and stop, or it may execute all the loops in an effort to find more than one solution. By executing all the loops Feenstra has detected several "bad" I.Q. questions that have more than one solution. If none of the four loops finds any solution, it is reasonable to include an additional output statement following the entire lot. The message it prints can vary according to taste; those who like to invest their programs with a little personality can have it print "Help!"

One can try out SE Q on questions from sample I.Q. tests found in several widely available books. In Eysenck's book there are eight complete I.Q. tests, which allegedly enable the reader to discover his or her own I.Q. The tests incorporate several different kinds of questions that appear on standard I.Q. tests, including questions that involve a missing number, a missing letter, a missing word, an odd-man-out, scrambled words and visual anal-

The new disk-based Tandy 600.

A breakthrough in portable computing power and versatility.

Finally—a portable that matches the performance of a desktop computer. Our new Tandy 600 features a 16-bit microprocessor, an 80-character by 16-line display, a built-in 3$^1/_2$″ disk drive that stores 360K of data and 32K RAM (expandable to 224K*).

Five Resident Applications

With Tandy 600's larger display and memory, Multiplan spreadsheet analysis can hold more information. Word processing is as easy as using MS-Word. You'll have quicker access to documents and better storage with the built-in disk drive. File is an electronic data base for names and addresses, expenses, client billing, inventory and more. And you can keep a large number of different files on the pocket size diskettes. With Telecom and the Tandy 600's built-in modem, you're able to communicate with other computers over phone lines. Telecom will even dial the phone number of anyone listed in the File program. Calendar helps you keep track of daily tasks and activities.

The easy-to-learn resident System Manager lets you run each application, exchange information between applications and manage the files created. You can even set it to turn itself on for unattended operations. And you can add BASIC language ROM (26-3904, $129.95) to write your own programs.

Tandy...Clearly Superior™

If you've been looking at portable computers, compare them with the Tandy 600 (26-3901, $1599). You'll be amazed at what it can do! Visit a Radio Shack Computer Center today.

Available at over 1200 Radio Shack Computer Centers and at participating Radio Shack stores and dealers.

Radio Shack®
COMPUTER CENTERS
A DIVISION OF TANDY CORPORATION

Send Me a Free Tandy 600 Brochure.

Radio Shack, Dept. 86-A-570
300 One Tandy Center
Fort Worth, TX 76102

Name________________
Company________________
____
Address________________
City________________
State________________
Zip________________
Phone________________

*RAM upgrades are in banks of 96K each (26-3910, $399.95). Prices apply at Radio Shack Computer Centers and participating stores and dealers. Multiplan and MS-Word/TM Microsoft Corp.

ogies [*see illustration below*]. A few subtypes are usually found within each major category. For example, in Eysenck's book there are three kinds of questions that ask for a missing number, namely the ordinary numerical-sequence problems I have already described and two other kinds typified by the following examples:

164	(225)	286
224	(——)	476

8	3	21
6	5	25
12	2	—

In each case the would-be genius must supply the missing number in accordance with some perceived rules. Feenstra's HI Q program handles such questions by procedures that draw on the same kinds of formulas as the sequence-completion program does. I encourage readers to try them; next month I shall give the answers to the above two problems, as well as to the problems posed in the illustrations below and on page 14.

Although HI Q answers only one major kind of I.Q. test question, the solutions to other kinds of questions can also be mechanized. In fact, a program that solves visual analogies was written more than 20 years ago by Thomas G. Evans as part of a Ph.D. dissertation done at the Massachusetts Institute of Technology. Heavy as it sounds, the essential ideas of Evans' program are easy to understand.

The visual analogies it solves all have the same form: figure *A* is to figure *B* as figure *C* is to one of, say, four figures listed as potential answers. The program selects the analogous figure by first determining a simple set of rules that can transform figure *A* into figure *B* [*see illustration on page 21*]. It then repeats the procedure with figure *C* and each of the four potential answers; in each case it generates a set of rules that can transform figure *C* into the potential answer. The figure obtained from the transformation rules that most closely resemble the rules for transforming figure *A* into figure *B* is selected as the solution.

Evans' program repeats essentially one operation five times. Two figures at a time, a source figure and a destination figure, serve as input. For each pair of figures the program then develops a three-part tabular description of how the source figure becomes the destination figure. First the program lists the spatial relations among the parts of the source figure; it then lists the spatial relations among the parts of the destination figure. Both descriptions consider only three spatial relations, *above, left* and *inside*. Finally, the program describes how parts of the source figure change into parts of the destination figure in one of four basic ways: each part can be altered in size, rotated, reflected or deleted.

Suppose figures *A, B* and *C* each have three parts, a circle, a square and a triangle. In figures *A* and *B* the program may label the triangles *a,* the squares *b* and the circles *c,* but it makes no attempt to label the parts of figure *C* in the same way. Instead it arbitrarily assigns the labels *x, y* and *z* to the three parts of figure *C.* It then develops its three-part tabular description for the pair of figures *A* and *B* and four more descriptions, one for each pairing of figure *C* with a potential solution. The last four tables all employ the labels *x, y* and *z* throughout.

The final operation of Evans' program is to make every possible substitution of *a, b* and *c* for *x, y* and *z.* Since *x, y* and *z* can be permuted in only six ways, there are six substitutions to be tried. One of the substitutions may convert the tabular description of the pair of figures *A* and *B* into the corresponding tabular description of figure *C* paired with one of the potential answers. The figure in this pair is the solution. Even if no perfect matches are found, however, the program can score the relative success of an analogy and so pick the substitution that yields the best match.

Patrick Henry Winston describes the visual-analogies program in his book *Artificial Intelligence.* Winston states that the program "works well," and he attributes its success to the use of an effective framework for representing knowledge about the geometric figures considered by the program. For example, instead of specifying how relations such as *above, left* and *inside* change from picture to picture, the program might have described how one figure in the first picture gets transformed into another figure in the second. Such a program might be extremely cumbersome, not to say ineffective, because it would have to check a much larger number of potential substitutions than Evans' program does. Indeed, the search for a good representation is a major theme of artificial intelligence: it is often the key that enables a computer to mimic some aspect of human problem solving.

By analogy the representation of objects in the mind is also the subject of much current discussion among cognitive scientists. In this context the study of artificial intelligence is often justified as an attempt to exhibit an "existence proof" for the mechanistic description of human capabilities. Thus, goes the argument, if a computer pro-

1. Insert the missing number.   3 7 16 35 ___

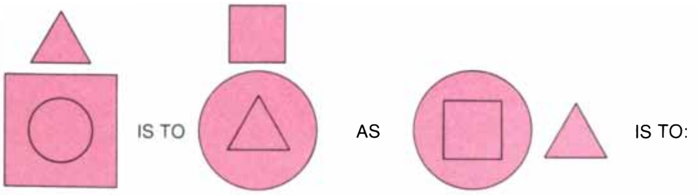
2. Insert the missing letter.   N Q L S J U ___

3. Insert the word that completes the first word and starts the second.   GRO (__) PER

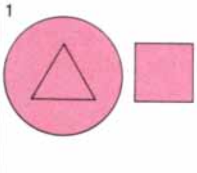
4. Underline the odd man out.
ANIMAL   ENGINE   IDENTITY   OCTAGON   UNICORN

5. Underline which of these towns is not in Italy.
NORLEFEC   DARDIM   SAIP   LIMNA

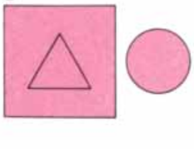
6. Which of the four numbered figures completes the top line?



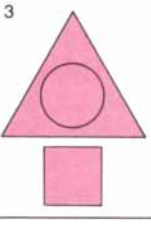
1



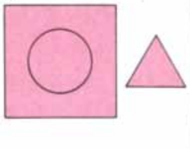
2



3



4



I.Q. minitest based on questions from Hans J. Eysenck's **Know Your Own I.Q.**

Vision
of
Grandeur.
rier-Lapos
Liqueur
PRODUCE OF FRANCE
BOTTLED IN FRANCE
Triple Orange
Triple Orange
Grand Marnier
Liqueur
PARIS — FRANCE
750 ml (25.4 fl.oz) - 80° Proof

gram can be made to simulate some aspect of human behavior, the representation of the behavior adopted by the program at least could serve as the underlying representation adopted by the brain. Nevertheless, it often seems that successful simulations of such behavior give little insight into how people do the same things. For all one knows there may be no relation whatever between the way Feenstra's or Evans' I.Q.-test programs perform and the way people solve the same kinds of problems. Presumably human intelligence deploys more general strategies in attacking particular problems.

This point brings me full circle to the reconsideration of human intelligence: what it is and how it is measured. As I have noted, Stephen Jay Gould has characterized I.Q. as a mismeasure of man. His criticisms carefully document two major fallacies that underlie the concept: the uncritical reification of an abstraction and the ranking of the reified abstraction along a single scale. Language itself accounts in large part for our tendency to make things of what are at best nebulous abstractions. Moreover, once we have persuaded ourselves that we are dealing with a thing, our reflex is to measure it.

In demanding a single numerical measure we succumb to the second fallacy, namely ranking. We want to reduce complex phenomena to a single scale. Such practices have led to excellent physics, but they have also led to some poor social science. I.Q. testing is a case in point; it is to the 20th century what craniometry was to the 19th. In both instances entire racial groups found themselves mismeasured not only because the measure was almost meaningless to begin with but also because there were biases introduced (either consciously or unconsciously) in the process of measuring.

Gould vigorously attacks biological determinism, the idea that human behavior is determined by genes, and he warns against viewing the capacities of our brains as direct products of natural selection. "Our brains are enormously complex computers," he writes. "If I install a much simpler computer to keep accounts in a factory, it can also perform many other, more complex tasks unrelated to its appointed role. These additional capacities are ineluctable consequences of structural design, not direct adaptations. Our vastly more complex organic computers were also built for reasons, but possess an almost terrifying array of additional capacities—including, I suspect, most of what makes us human."

With this last metaphor Gould has put his finger on what I find most unsettling about a relatively simple computer program that can score at the genius level on an I.Q. test. Does the score on the test measure the intelligence of the computer? If it does not, just how does one go about measuring the intelligence of a computer, whether it is made of silicon and plastic or of carbon and tissue? The answer: Probably not by running some I.Q. program through a battery of tests.

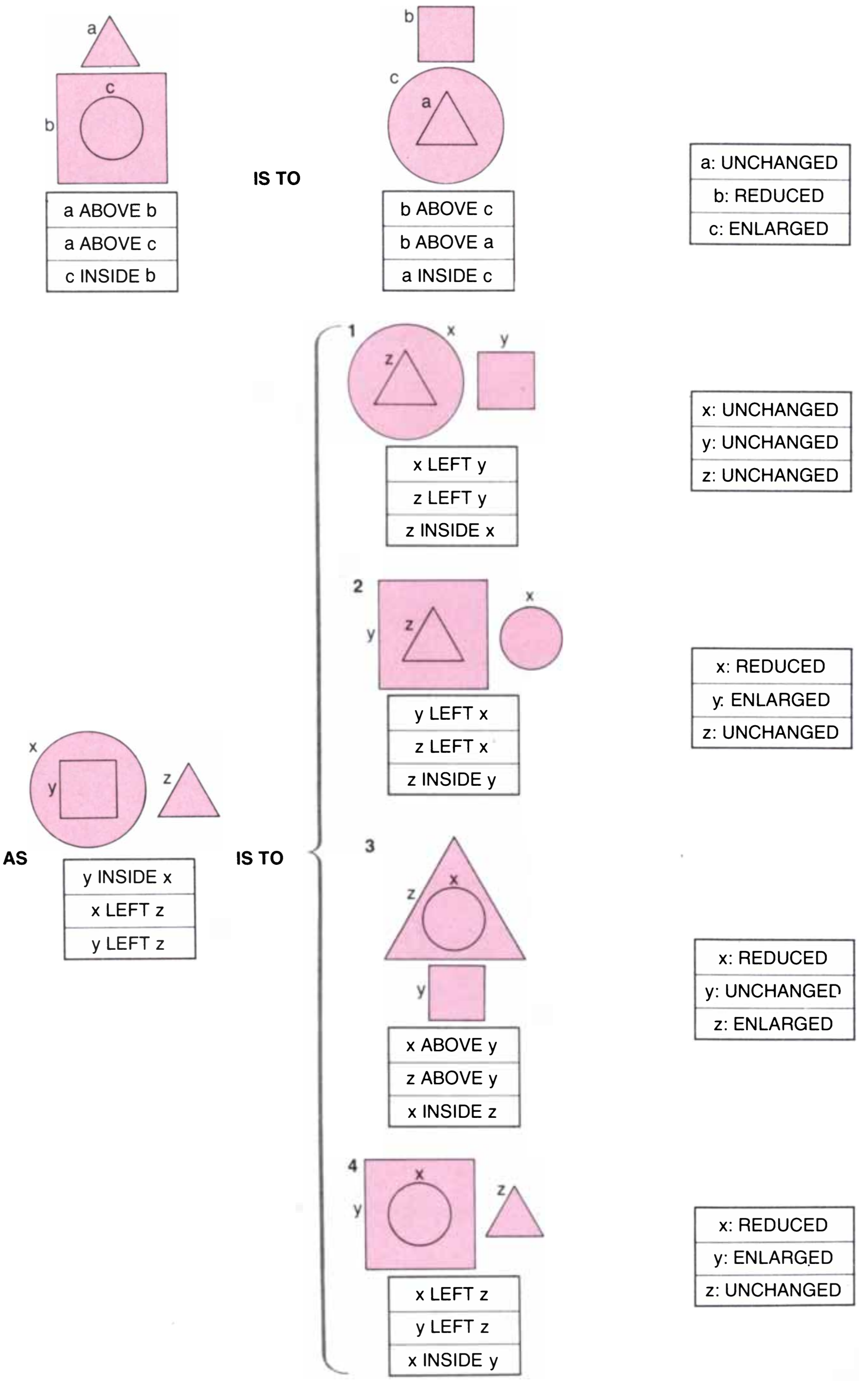
Golomb rulers, the subject of last December's column, turned out to be the toughest project that readers have yet faced. Many were called but few were chosen, so to speak. Several readers even sought to claim a $100 prize offered by the inventor of the rulers, Solomon W. Golomb of the University of Southern California.

A Golomb ruler with *n* marks is the shortest ruler possible with the following properties: it bears *n* distinct marks (including the endpoints) at integer positions, and it measures as many integral lengths as possible from 1 to the length of the ruler, each length in at most one way. A distance can be measured by the ruler only if it is the distance between some pair of marks. If the same distance can be measured between more than one pair of marks, the ruler is not a Golomb ruler.

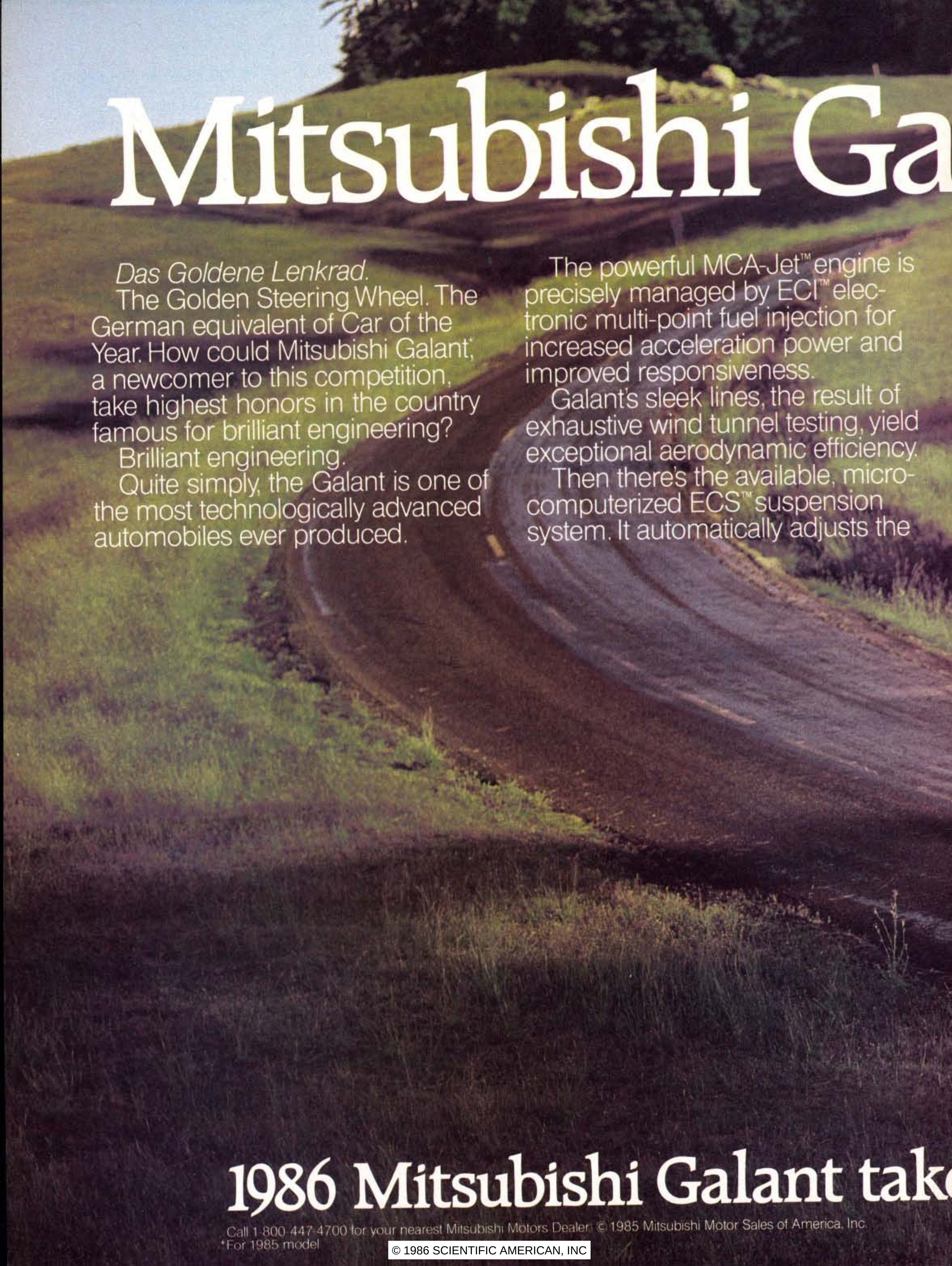
At the time the December column was published, no Golomb rulers were known with more than 13 marks, and the shortest ruler known with 15 marks was 155 units long. Soon thereafter Douglas S. Robertson of the National Oceanic and Atmospheric Administration discovered a shorter 15-mark ruler only 153 units long. Then during the Christmas holidays James B. Shearer of the IBM Thomas J. Watson Research Center programmed an idle computer to search exhaustively for rulers, and the computer has now turned up Golomb rulers with 14 and 15 marks. The 14-mark Golomb ruler is 127 units long and has marks at 0, 5, 28, 38, 41, 49, 50, 68, 75, 92, 107, 121, 123 and 127. The 15-mark Golomb ruler is 151 units long and has marks at 0, 6, 7, 15, 28, 40, 51, 75, 89, 92, 94, 121, 131, 147 and 151. Shearer writes that he saved much computing time by assuming the middle mark on the ruler is to the left of the geometric middle.

Another problem posed by Golomb has generated the claims for the $100 prize. The claims made so far are invalid, apparently because they are based on misunderstandings of the problem. Golomb has urged me to clarify matters by restating it. Find two different rulers (whether of minimal length or not), each having the same number of marks for some number greater than 6, that measure the same set of distances; again, no distance on either ruler can be measured between more than one pair of marks. Reflections, such as the ruler with marks at 0, 2, 5, 6 and the ruler with marks at 0, 1, 4, 6, are not counted as different. There are infinitely many known pairs of rulers, almost all of them nonminimal, that solve the analogue of Golomb's problem for six marks. For example, one such pair have marks respectively at 0, 1, 4, 10, 12, 17 and at 0, 1, 8, 11, 13, 17. They are nonreflecting, essentially different rulers, but they both measure all distances between 1 and 17 except 14 and 15. The prize will go to the first person who discovers such a pair of rulers with more than six marks each.

The advisory network to help programming novices with projects stemming from this department has run into unforeseen difficulties: there are hundreds of advisers but almost no tyros. The name tyro may have been ill-advised. Has it put people off who program with little success? It is time to send me a card bearing your name, address and telephone number, in care of this magazine.



The solution of visual analogies in Thomas G. Evans' dissertation



Mitsubishi Ga

Das Goldene Lenkrad.

The Golden Steering Wheel. The German equivalent of Car of the Year. How could Mitsubishi Galant*, a newcomer to this competition, take highest honors in the country famous for brilliant engineering?

Brilliant engineering.

Quite simply, the Galant is one of the most technologically advanced automobiles ever produced.

The powerful MCA-Jet™ engine is precisely managed by ECI™ electronic multi-point fuel injection for increased acceleration power and improved responsiveness.

Galant's sleek lines, the result of exhaustive wind tunnel testing, yield exceptional aerodynamic efficiency.

Then there's the available, microcomputerized ECS™ suspension system. It automatically adjusts the

1986 Mitsubishi Galant tak

Call 1-800-447-4700 for your nearest Mitsubishi Motors Dealer. © 1985 Mitsubishi Motor Sales of America, Inc.
*For 1985 model

lant. Natürlich.

suspension firmness to suit road conditions. It keeps the car level regardless of load. And lowers the ride height at highway speeds for more handling stability as well as reduced wind resistance.

Innovative ergonomic design provides sumptuous comfort. Even the rear seat reclines. And one of the 11 functions of the ETACS IV™ electronic control system prevents you from locking your keys in the ignition.

An optional stereo even offers you steering wheel-mounted switches.

The list of its wondrous features goes on. And on. So test drive a new Mitsubishi Galant. You'll quickly understand why this automobile is recognized as perhaps the world's most ingeniously engineered luxury sedan.

In any language.

MITSUBISHI
MOTORS

s you where you want to be.™

BOOKS

A toolbox of musical instruments and a centennial celebration of Niels Bohr

by Philip Morrison

HANDBOOK OF INSTRUMENTATION, by Andrew Stiller. Illustrations by James Stamos. University of California Press ($65). Hefty and typographically handsome, this volume is a startlingly comprehensive field guide to the living instruments of music. The author, a composer at the State University of New York at Buffalo, plainly a fond and tireless investigator, and his artist collaborator have sought out and assayed the nature and behavior of every instrument "currently in use for the performance of classical and popular music in North America." A couple of hundred instruments are drawn to careful scale in meticulous line. The washboard is no less carefully rendered than the intricate keys of the heckelphone (a tenor oboe) or the strings and frame of the grand piano. This census is current and documented; it is not historical. Folk and non-Western instruments are excluded unless they have crept into wider use.

Nine-tenths of the instruments presented can be heard anywhere on the planet that listeners might experience classical or popular works by such composers or improvisers as Domenico Scarlatti, John Cage or the Grateful Dead. Exemplary compositions are cited for each instrument. The intent is not at all that of a dictionary; it is to make the many sources of living music into familiar tools for those who would compose music today. Mere listeners and players may of course eavesdrop; what they learn may not be particularly useful, but it is delightful in both depth and breadth. Builders and buyers of instruments will also find much of interest.

The instruments are grouped chapter by chapter into the familiar families of the orchestra. The availability of adept players is estimated for every device, on a scale that runs from ubiquitous to very rare. Specific notation, a brief historical and physical analysis of function and limitations, the vital statistics of range and loudness, and an earwitness account of the skilled player's entire battery of artifices form the entries. Among the appended material there is even a list of a manufacturer or two for each instrument.

The wildly varied percussion family is treated in detail. Timpani, vibraslap, maraca and thunder sheet as well as all the rest, along with the brushes, sticks and mallets they severally demand, require a third of all the space given to modern instruments. A second, shorter part treats contemporary reconstructions of instruments from the past, all those recorders, krummhorns, sackbuts, viols, harpsichords and lutes that again, often after centuries of neglect, pleasure the enthusiasts.

Summarizing a book whose strength is in its comprehensiveness is not of much value, but a few tasty kernels can be displayed. The voice is treated for its unequaled variety. Most of the sounds we can make find use in one or another natural language. The International Phonetic Alphabet is included, rather out of date for linguists nowadays, inadequate for clicks and plosives and Donald Duck noises, but still a "godsend to the composer."

The registers of the untrained voice are surveyed, with a little helpful physics. The tension of the vocal cords fixes the pitch range; the cavity resonances of throat and mouth determine timbre. Four states of tension are available to the cords. They are called the registers; they differ both in pitch range and in harmonic content. The four are the whistle register, falsetto (or head), chest (normal speech) and growl. Classical voice training teaches the elision of most of the differences. The whistle register, thin and piping, is heard mainly in the shrieks of children at play. It has almost never been used in music; an exception is Mozart's Queen of the Night. Mixed with falsetto, it evokes "a spectacular kind of insanity." The growl register is performed more easily on the inhale; no lower limit on its pitch is listed. There is a table of differences between speech and song, to clarify such hybrids as *sprechstimme.* Special voice effects are produced by the buzzy kazoo, by the mailing tube, whose resonance obscures the vowels and so dominates a song, by the bullhorn and even by the gallon jug. This section ends with a wistful note on whistling, a "great unexplored region" in music. Perhaps there are "one or two people in the world who can whistle and hum two different musical lines simultaneously."

To the account of so important a topic affix a sample of unexpected minor points. Snare drums claim a few pages; they are the most Western of all drums. The two drumheads are dissonantly tuned, and wire coils are stretched across the lower head. Those coils rattle as the drum is beaten, and the result is an almost unbroken band of white noise. The sledgehammer (cited from a piece of Alban Berg's) is used only on a massive slab of wood or metal kept on the floor. It takes two hands (give the percussionist room!) and is hardly to be swung very fast. The double-reed rackett "is surely one of the strangest creations of the human brain." Within the squat cylinder of the instrument's Baroque version the bore is folded on itself 10 times; many little tone holes are arranged in a loony fingering system for the nimble player. It sounds like a bassoon of the period.

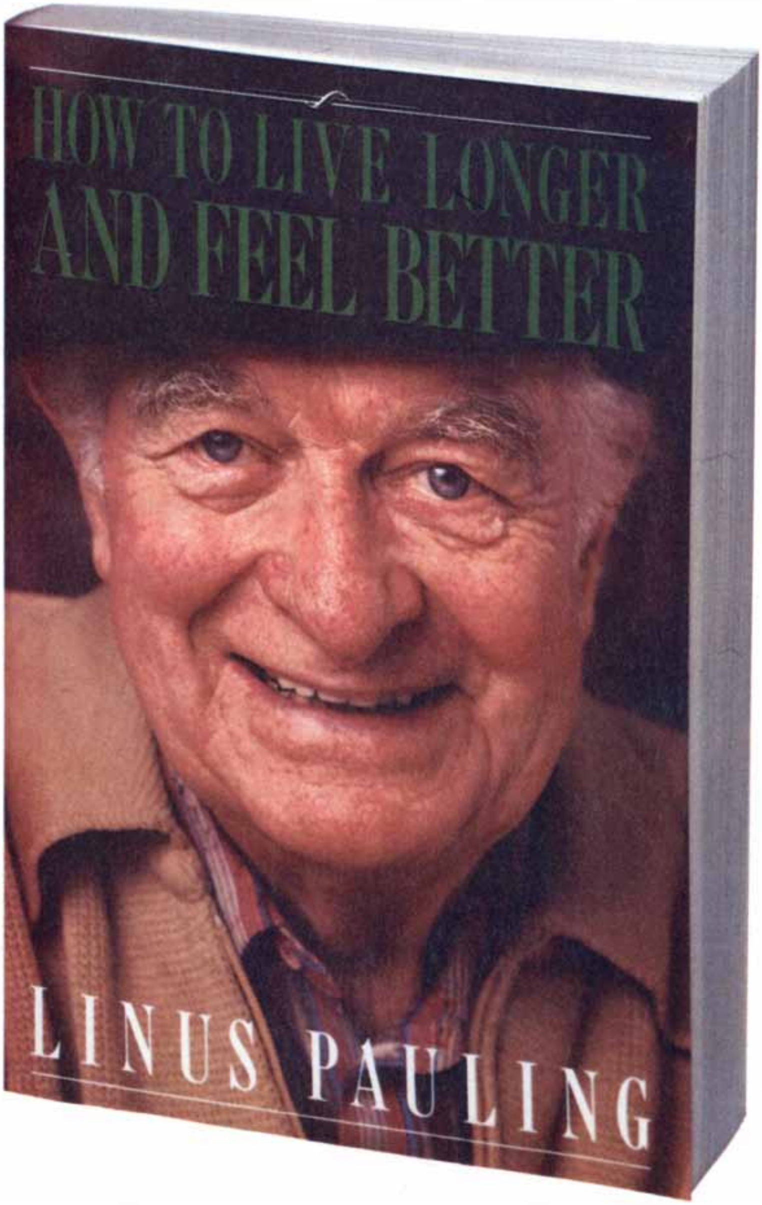
Live electronic music is well treated. Its young audiences must now be the largest of them all. The "coarse and powerful" electric guitar in its several versions remains central, replete with fuzz, reverb, vibrato, wawa and phaser. The signal source consists of two or three magnetic pickups that take their input directly from metal strings; there is no sound box in an electric guitar. "Its genius is for melody, pure and simple." What you hear is a much amplified loudspeaker output, after intermediate processing to choice. The fuzz box produces a distortion that can make the second harmonic louder than the fundamental; it evokes the genuinely overdriven amplifiers of the past. Wawa is a low-pass filter with a pedal-controlled cutoff. Reverb and vibrato are approximated by filter effects; the phaser is a kind of comb filter that eliminates by means of delayed recombination harmonics that are mainly opposed in phase. East and West can unite around the rare electric sitar.

The veteran Moog synthesizer is a modular organlike console sporting a thicket of patch cords. It has half a dozen standardized tunable oscillators, each capable of use as control voltages for other oscillators as well as for direct audio output. Thus they yield sawtooths, square waves, pulses, ramps and triangular waveforms along a parametric continuum that

*If you plan on living to be 76,
why stop feeling good at 40?*

JUST PUBLISHED

HOW TO LIVE LONGER AND FEEL BETTER



LINUS PAULING,
LINUS PAULING INSTITUTE
OF SCIENCE AND MEDICINE

In HOW TO LIVE LONGER AND FEEL BETTER, two-time Nobel Laureate Linus Pauling proves you don't have to be a ''health nut'' in order to add years of feeling good to your life. His inexpensive and uncomplicated regimen includes . . .

- three meals a day at regular times; no between-meal snacks, which are usually very high in sugar
- moderate exercise two or three times a week
- six to ten hours of sleep each night
- moderate alcohol consumption (two or three drinks a day)
- optimum supplementary vitamin intake daily.

In addition to showing how vitamins work, Pauling discusses the role of proper nutrition in strengthening the immune system. He explains what each of us can do about doctors who are good at treating some illnesses, but not so good at helping us to prevent them. And, armed with conclusive experimental and clinical evidence in support of his ideas, he exposes the true facts behind the Mayo Clinic's distortion of his successful ortho-molecular treatments of terminally ill cancer patients.

1986 352 pages
$7.95 ISBN: 0-7167-1775-1 paper

Available at fine bookstores or send $7.95, plus $1.50 shipping, to:

W.H. Freeman and Company
4419 West 1980 South
Salt Lake City, UT 84104

is under knob control. White- and pink-noise generators ("a bit like distant thunder") add their randomized nuances, and envelope controls modify attack and decay of the waveforms through flexible, voltage-controlled attenuators. Nonlinear mixing, scrambling and clipping are frequently applied as well. With tape recording of such shenanigans, the distinction between composition and live performance has faded away.

Of course all is change in the world of circuitry. Digital synthesizers under microprocessor control are in growing use. So far they have mainly generated discrete pseudomorphs of the analog waveforms pioneered by the Moog. They pay their dividends in instrument compactness and cost.

Yet chiefly it is the acoustic universe of boxes and membranes and tubes and strings that is celebrated here in an astonishing variety of scale, material and form. These artifacts and practices, at times beautiful, at times absurd, manage to couple more tightly to the ancient art they all serve than they do even to the pulsing air itself.

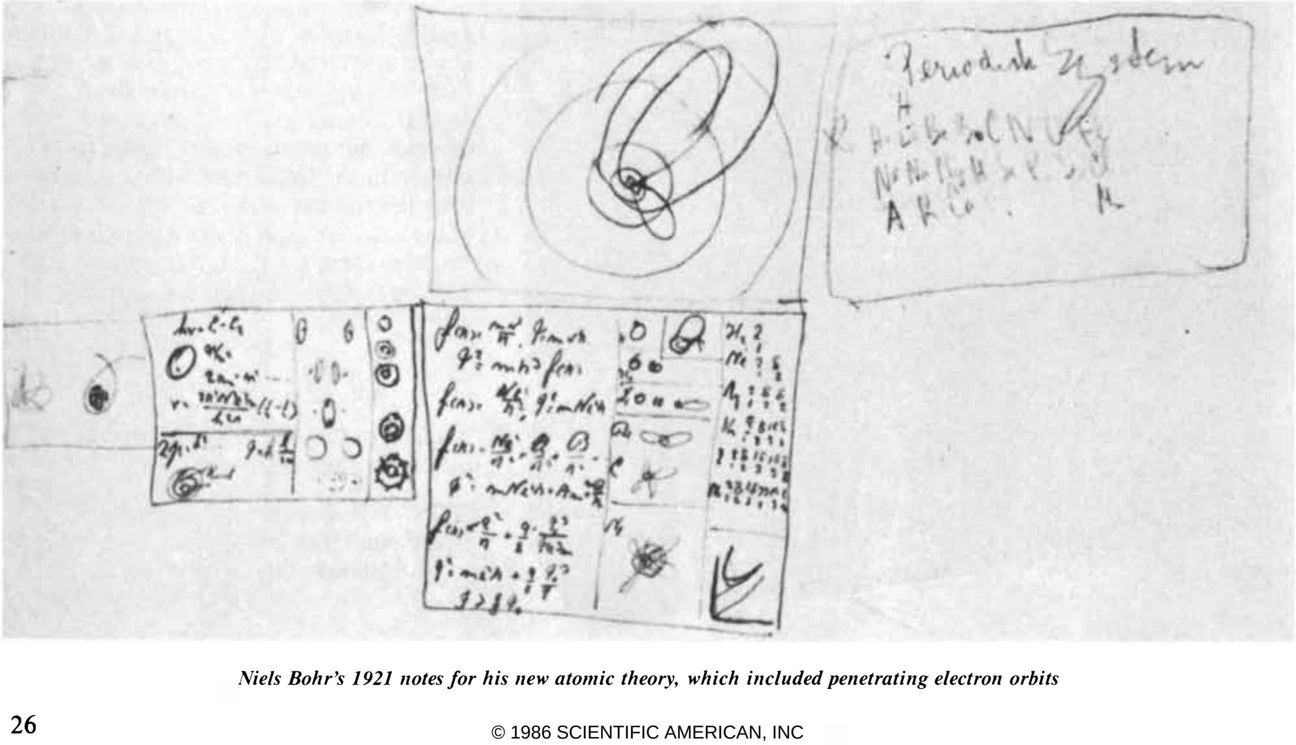
NIELS BOHR: A CENTENARY VOLUME, edited by A. P. French and P. J. Kennedy. Harvard University Press ($27.50). "Dear Dad," wrote W. L. Bragg, then a physics student at Cambridge, to an older physicist: "I'm so glad you liked the notes on Jeans.... I got an awful lot from a Dane who had seen me asking Jeans questions, and ... came up and talked over the whole thing. He was awfully sound on it, and most interesting." The Dane was Niels Bohr at 26, a postdoc in the Cavendish, soon to make his way to Rutherford at Manchester, just as the nuclear atom was born out of the closely watched scattering of alpha particles. Bohr's own tussle with the ordered spectral pattern of the atom began then and there, in about 1911.

Long before, in 1885, J. J. Balmer had published his remarkable integer formula for the visible spectrum of atomic hydrogen. Bohr was born in that same year; it took physics three decades to explain those small integers in optical spectra. Bohr's atom model was crude and remained seriously incomplete, but it was absolutely on the right track. When he made the first easy extension to the spectrum of helium, interest quickened. In October, 1913, he refined his mechanics a little to take account of the motion of the nucleus itself, since of course the sunlike nucleus did not infinitely outweigh its planet electrons. It moved too. The outcome was a soon-confirmed fit to the measured line positions of helium accurate at the level of a couple of parts in 100,000. George de Hevesy, a close friend, wrote Bohr (in an orthography all his own) of Einstein's reaction to the news: "When he heard this, he was extremely astonished and told me: 'Than the frequency of the light does not depand at all on the frequency of the electron ... And this is an *enormous achiewement*. The theory of Bohr must be then wright."

Of course it was not right. It resembles the atom no more than a quick pencil sketch resembles a living face, as O. R. Frisch put it. The real Jacob, quantum mechanics, appeared only in the mid-twenties. How much Bohr meant to this second coming, even more as a personal focus of understanding than in published papers, is the next high theme of this volume, glowing with affectionate recollections by half a dozen eminent physicists of the Copenhagen days between the wars. A warmth born of community in the long struggle for understanding is radiated from every reminiscence of Bohr. He was a friend in decisive action no less than in spirit. It is moving to read of his prompt tour of the German universities during the fearful year of 1933 once the Third Reich rose in fire; as he went from city to city and physicist to physicist, he quietly set up "the lines of escape."

The book contains Bohr's published masterpiece, that long review, half history, half reanalysis, of his decades of dialogue with Einstein over whether quantum mechanics can be a complete theory or is only a statistical shadow of a more causal but more hidden reality. He wrote it for Einstein's 70th birthday in 1949. The depth of the issue has kept it alive. Two or three shorter papers summarize and discuss the present-day status of arrangements that allow the experimental choice between two interfering measurables to be delayed until the significant interaction seems long ended. Just as Bohr contended, the new results fulfill the quantum forecasts based on the indivisibility of experiment, but they surely contradict less subtle versions of physical reality.

The third great theme in Bohr's life



Niels Bohr's 1921 notes for his new atomic theory, which included penetrating electron orbits sounds yet in all our lives. It is fission. That nuclear phenomenon entered the world with the year 1939. It was Bohr who first brought across the Atlantic in person the word of how Frisch and Lise Meitner had explained and confirmed the enigmatic report of the radiochemists that uranium could absorb a neutron and then break into two heavy fragments with an unprecedented release of energy. Bohr and John Wheeler spent six months clarifying the physics of the process, in particular which isotopes would show fission. Their decisive fission paper appeared just as grim war broke out in Europe. It is reproduced here in part, along with a chronicle of the people and the events of that fateful year for physics and the world.

Bohr faced the consequences too. First he escaped from Denmark in the bomb bay of a fast Mosquito, converted so that B.O.A.C. could fly civilians from neutral Sweden to Britain at altitudes above the Luftwaffe's reach. He made his way first to Los Alamos and then soon to the side of Churchill and Roosevelt. The memorandums he prepared for the two leaders are here, again in a mix of anthology and new historical comment. His plea for an open world around nuclear energy release was not heard, during the war or afterward. The P.M. at once took a dislike to that man "with his hair all over his head" and even grumbled that "Bohr ought to be confined" as a"great advocate of publicity."

A section learnedly treats Bohr's philosophy, again by means of old papers and new. It includes accounts of the Copenhagen views both as seen from a base in Sanskrit dialectic and as debated in the U.S.S.R. by the thinkers of Marxism-Leninism. (Hideki Yukawa is cited as observing that in Japan Bohr's complementarity looked quite evident: "You see, we ... have not been corrupted by Aristotle.")

The book is lighthearted in its admiration, and it does not omit the times when Bohr fell into error: that bad year before the discovery of the Compton effect when he despaired of the conservation laws, or the occasion when he guessed that success in molecular biology would require complementarity at some deep functional level. Light was not the key to life. That turned out, as far as we now see it, to be more closely related to Schrödinger's notion of the aperiodic crystal. But it is striking to read H. B. G. Casimir's remark that it was Bohr who steered him toward the zero-point energy of the photon field as a description of his wonderful derivation of the forces between closely spaced conductors. Bohr seems rather casually to have thought decades ahead, exploring a major puzzle alive in the sophisticated present.

A few pages in context, picture and verse, are reproduced out of the 1932 Copenhagen parody of *Faust,* written essentially by Max Delbrück. The parody's Mephistopheles was clearly meant to be Wolfgang Pauli; the Lord God rather resembled Bohr in a top hat. The entire script was published by George Gamow in his autobiographical *Thirty Years That Shook Physics.*

Plainly an elderly physicist can enjoy the volume, with its cargo of apt figures and citations. About a fifth of the text is given to Bohr's own papers, another fifth to reprints of older studies; the rest has been written for the centenary year. Every student of physics will want some access to these instructive matters, access made so easy here that the volume is a fine buy for every library such students frequent, from high schools on up. The general scientific reader will also find much of interest in this diverse and well-presented collection of 40 papers. (Observe how useful are the notes the editors have provided, as their final admirable contribution to the welfare of the International Commission on Physics Education, which will benefit from the royalties on the book.)

BLACK CARBON IN THE ENVIRONMENT: PROPERTIES AND DISTRIBUTION, by Edward D. Goldberg. John Wiley & Sons, Inc. ($29.95). The stuff is everywhere. What we know of it is inferential at best; even a tight definition seems premature. Those black complex particulates exist at micron scale, give or take a factor of 10. A typical particle is impure: a working majority of its atoms are always carbon, but hydrogen, oxygen, nitrogen and sulfur may also be prominent. Carbon particulates are the result of incomplete combustion. Such is the matter examined in this timely monograph.

Nowadays we learn of black carbon copious worldwide in the clay layer that anciently sealed the Cretaceous, when somehow the stars threatened the dinosaurs; of late we quite reasonably fear a second coming in black wrath on our own heads. Edward D. Goldberg, a chemical oceanographer at the Scripps Institution of Oceanography, prepared his compact and varied survey out of a scattered literature still in rapid growth, sometimes resting an entire chapter on a few sooty papers. In no way final, its rich data, here and there confused or even inconsistent, present a view of these unexpected processes that is bound to be vivid and surprising to any reader who has a head for figures.

Attend to the auditors of the world carbon accounts. The author and his colleagues measured fluxes of black carbon at sea in the coastal waters of the Pacific and on the open ocean as well. No doubt the material is born on land; the open oceans receive only a thousandth of what falls on the continents and on the sea surface near land. But the measured falls confirm remarkably well the estimates of biomass burning in all its variety that reckon from total areas of vegetation types, biomass densities and efficiency of burning. It appears that during the decades 1960–80 roughly equal carbon-black contributions were made globally by the clearing fires of slash-and-burn tropical agriculture and by the burning of farm and timberland wastes. Smaller but comparable inputs came from both the wild and the deliberate fires in savanna and bushland, and from wood fuels in households and industry.

On greater Los Angeles the soot rains thick, the bulk of it from all those heavy diesel trucks, whose yield of black is tenfold their proportionate consumption of fossil fuel. Yet the flux density of carbon black on that restless engine-loud basin is only 10 times the density found over lonely coastal seas. A calculation concludes that all the burning of fossil fuels is by two orders of magnitude a negligible global source of the black compared with biomass fires widespread over farms, forests and grasslands. The black country is grim but its area is minor. World satellite photographs recorded at night show farmlands burning more widely than the lights of all the bright cities, and the bush fires are much sootier than most of those fueled by oil and powdered coal.

What is the carbon-black particle? Characterization is no simple matter, even though analytical methods are growing in power, from the electron microscope to the Raman spectrum and nuclear magnetic resonance. One submicron soot species, formed by burning hexane, is described in molecular detail. A rough idea of this object, not claimed as typical, might be that it is a porous complex of carelessly stacked, tiny rafts of hexagonally packed hydrocarbon rings. Each raft consists of a few score such molecules, much infiltrated by other less regular and less carbonaceous rings; the interstices host all kinds of open oxygenated carbon chains and folds. Crystalline precision is achieved only rarely. The resins of biomass burn to such soots, whereas the woody portions more likely yield soots of elongated, even cellular form, the ruins of cells.

The charcoal accounts by no means balance. The annual fallout of carbon black would amount to the total carbon of the biosphere in a geologic flash of less than 100,000 years' duration. The rate is probably too fast to accept even if there is a grand hidden methane input from the depths. The fire-born stuff is notoriously insoluble and refractory. Yet something surely eats the surface-rich tiny carbon-black particles. There are signs of such processes, both of inorganic breakdowns by photochemical reaction and by slow oxidation and of microbial degradation. One tracer study did show that soils full of healthy microorganisms generated carbon dioxide from labeled carbon blacks much more rapidly than sterilized samples of the same soils. But at best the changes are slow on the laboratory time scale, taking decades under realistic conditions. Their study is hence tedious and uncertain.

The record in recent sediments is not yet entirely clear. The charcoal content of bottom samples from Lake Michigan is dominated for a century before 1900 by the small particles from the burning of biomass. Then the count leaps upward and the particles, now coal soot, grow in size. A maximum comes in about 1960. It seems clear that the lake received increasing fallout from the smokestacks of growing Chicagoland until stack controls came in. The ooze at the bottom of the Lake of Clouds among the green forests of Minnesota records a steady charcoal influx, marked by meaningful fire peaks, over a period that extends some 1,000 years into the past. In Pacific deep-sea cores the record is more than 50 million years old. Here too there is a steady rise of about a hundredfold in charcoal input during that time. It reaches a plateau within the past few million years; it does not markedly reflect the coming of humanity at all. Could that rise be a consequence of wildfire in the grasses that spread over the Asian and American steppe during a million summers?

The book closes with an interesting summary of analytical methods and a very diverse set of references. Few geochemical volumes can be so closely linked to the growth of human society. All the same, the strongest inference is that fire on a large scale is much older than our own lineage. That Promethean gift, now potentially thermonuclear in scale, seems not yet quite safe in our hominid hands.

COMMUNICATION AND NONCOMMUNICATION BY CEPHALOPODS, by Martin Moynihan. Indiana University Press ($32.50). A decade ago these columns noted an unusual small book on the behavior of the New World primates by Martin Moynihan, who worked then as now at the Smithsonian Tropical Research Institute. That book was based on his 17 years of observation of our quick and clever little cousins in field and laboratory.

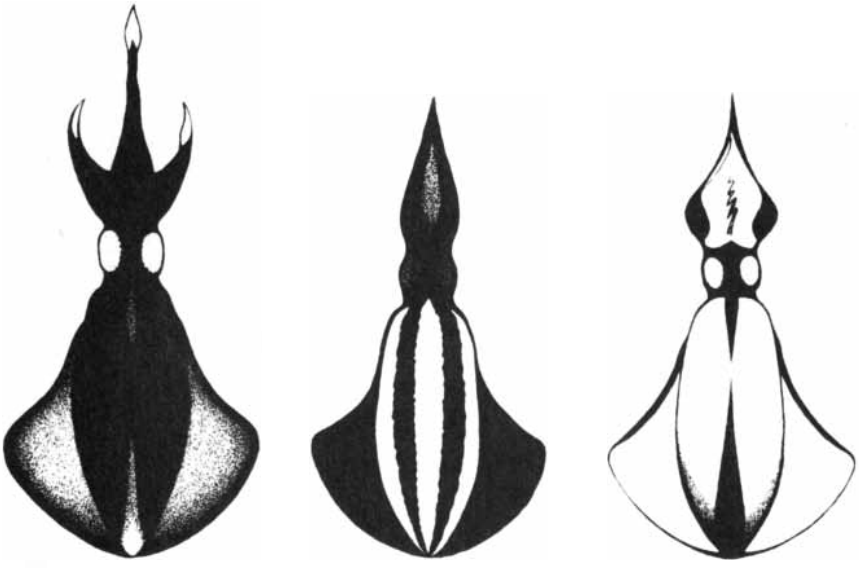
What Dr. Moynihan described in its pages was largely a search for language, arguably present among marmosets and titis through a wide variety of social and even sequential squeaks, trills and grunts, filled out with many gestures and visual displays. All of this was simply and evocatively drawn by the author. He was personal, theoretical and almost quizzical in his conjectures and reflections. The conclusion was sharp enough: most of the patterns of language are there, but its essence may be absent. The patterns are innate and unvarying. Monkey utterances are not learned. Nor do the active creatures show any sign of the generation of new combinations so plain, so indispensable in human speech and even in gesture, the linguistic productivity on which human society rests. A hint of new postures and movements was seen for a few individuals under social stress. Their fellows ignored these quirks, although Moynihan could often learn to judge the state of the animals from their unusual behavior.

It was natural then to refer to the author as a primatologist. Actually that inference was too narrow, since at about the time he wrote the earlier book he had left the green forest canopy to work in the blue-green offshore waters of the Caribbean reefs. There he has been watching the cephalopods, in particular a gregarious species of smallish squid.

Once again Moynihan has prepared out of his field work and his reflections a personal, quizzical, breezy and theoretically sophisticated small book concerning a specific form of animal communication. The essential information is presented chiefly in a few dozen black-and-white wash drawings from life and from the literature on the coleoid cephalopods (those with a dwindled internal shell), mainly the octopus, the cuttlefish and the squid, all of warm shallow seas. The images are so striking in their bold contrasts amidst fluid forms that a reader might easily think he had opened a study of a trendy new school of art. But it is no human art; it is the mime theater of the living coleoids, the subtlest of all invertebrates, members of a lineage whose links of descent go back to a time before the coming of mammals.

These marine creatures signal in silence. They do not react to sounds either in the laboratory or in the field. Perhaps they are thus protected from the powerful sonar pulses of their cetacean predators. They see well, although as the data now stand coleoids are said to be color-blind. A unique skin structure of a myriad dye sacs, each sac capable of voluntary change from pinhead invisibility to full expansion as a patch of color, allows color changes in complex and controlled patterns to flash over the entire body surface within a fraction of a second; only fireflies and butterflies can change their appearance with anything like the verve and assurance of coleoids, and the insects' options are stereotyped in comparison. The painterly changes and a rapid control of coleoid body texture itself along with its color pattern are unmatched. They can also give off into the water small clouds of pigmented fluid, the ink of the squid, dark or light.

Ritualized, exaggerated pattern display is common. Some three dozen



Different coloration patterns of the squid **Sepioteuthis lessoniana** patterns have been recognized among both squid and octopus, listed here from Colorless to Dymantic (eyelike) Spots and Zebra Stripes. The drawings are persuasive. One shows a squid pulling arms and head deep into the mantle cavity, until the entire beast is shrunk to the form of a ball, and on the ball is flashed up a spotted man-in-the-moon face. "The models could only be vertebrate." Another shows young squid in Dark color habit, their arms contorted and upward curled, very like their own ink clouds, earlier discharged and floating slowly by. "The animals were, in fact, mimicking their own decoys."

Ritual patterns are as often antidisplays as they are displays. They are meant to be seen, but to mislead and startle. Quick alternating sequences are frequent, usually directed at predators. These are baffle performances. The same patterns may be used toward other squid as part of sexual or social contact, but they do not then show the stereotyped sequences recorded in the baffle. "They can be trickier still." An entire group may alternate as one between Bars and Streaks as predators draw near. Yet a single performer is often firmly out of step, each time flashing the opposite. Is it trying to look different? "If so, it is reacting to its surroundings in a more . . . sophisticated way than even the most expert practitioner of crypsis."

"Cleverness is not all." Coleoid visual vocabularies are complex and organized. Pale seems correlated with fear, silver flashes with sex, for instance. There are clear signs of some kind of self-awareness: they must swiftly perceive both their surroundings and themselves—anyone who has seen coleoid performance in film can concur. Moynihan detects elements of a simple reef-squid grammar. There are strong signifiers, such as an outspread body form, and three weaker categories. Side Stripes or positional postures such as Head-down serve as modifiers. A standard background color suffuses the animal when it is undisturbed. Unritualized performances are common; they appear to serve social purposes in the gregarious forms. Here species differ widely. One has no Basic background; it flickers constantly, changing tone at the slightest provocation. Again there is no evidence for those new strings of old symbols humans so easily utter.

Maybe it will yet appear. We need plenty of motion pictures, with both signalers and responders on camera. That has never been done. The common octopus, a superb mime that carries the authority of a self-conscious artist to the untrained film watcher, is too solitary an animal. Signaling experiments and imitation to elicit response are both tough. "Not everyone can blush in streaks or bars." It will be a while before we can hope to recognize in that dazzling costume theater of cephalopod form the oldest and most alien script the earth may hold.

THE PICTURE BOOK OF QUANTUM MECHANICS, by Siegmund Brandt and Hans Dieter Dahmen. John Wiley & Sons, Inc. ($29.95). Look into a mirror. The image, those watchful eyes themselves, the source of light—all are parts of an unchanging system in steady state. Mainly we think of the situation not only as steady but also as in fact static, involving time no more than do the geometric diagrams bristling with rays that are usually used to explain it. Of course we can do better. Surely light pours out of the source, some portion of it to flow in turn to face, mirror and eye, if at indecent speed. Out of such an account the mind has a chance to unpack the causal chain, to imagine the successive interactions that physically partition and redirect the light. The steady-state abstraction is much simpler, but it offers a less satisfying intuitive grasp than the step-by-step narrative in time does.

Two physicists at the University of Siegen in West Germany have filled their attractive text with such time sequences in computer-generated diagrams and with quantitative contour maps of apt solutions to key equations. Their aim is the presentation of the "principal ideas of wave mechanics" in such a visual way that students can build a quantum intuition out of their graphics, much as we all do for classical physics from the flight of balls or from wave patterns realized in the ripple tank.

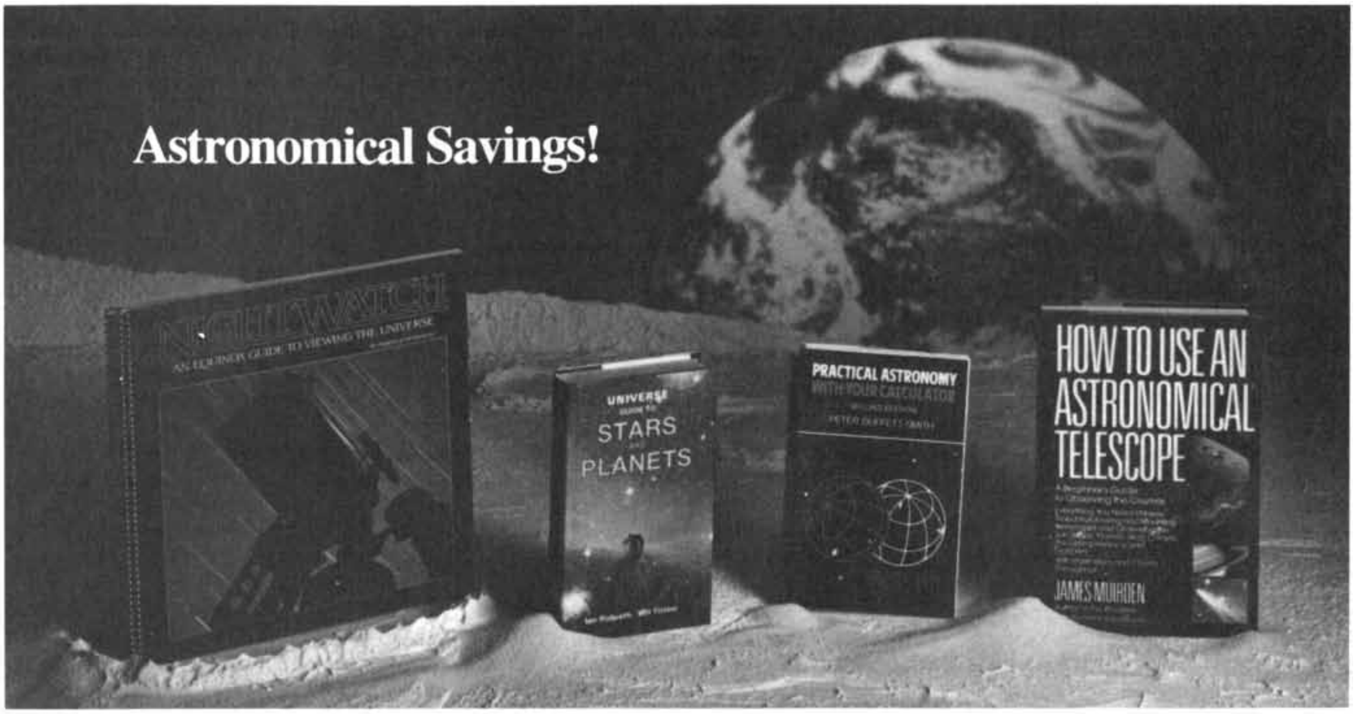
Nine of the 14 chapters center on the strict causal development in time that wave mechanics presents even for phenomena that in the end must be realized in the discrete and chancy world of photons and atoms. It is instructive to see the march of a packet of light falling on a glass surface as the event is represented in half a dozen carefully placed and scaled graphs created by the authors' interactive program. The fields first reflected interfere with the still-oncoming tail end of the pulse, to create a wiggly structure that soon enough settles into two distinct parts: the small, smooth pulse coming back while the slowed and compressed main signal moves on into the glass.

A few chapters later we are ready for a quite different yet recognizably related sequence in time. A Schrödinger wave packet represents a particle bound within two rigid walls, all motion limited to the single dimension of a line. The particle pulse, which begins by bouncing back and forth between the walls, slowly spreads. Each wall acts the way the mirror surface did for light, so that interference patterns first appear in turn at the two walls, until the packet has widened enough to show the effect near both walls at once. The fundamental role of the spectrum of individual energy states, whose superposition represents the particle in the well, is soon manifest.

Several chapters treat the motion of two coupled oscillators in one dimension. Not easy to puzzle out, but rewarding, these contour maps display joint probability distributions shifting in close analogy with the familiar classical effect of the rhythmic transfer of energy back and forth between the interacting pair of particles. Here a striking extension is made to the additional quantum interference effects that arise when the two oscillators are indistinguishable, namely the bosons and fermions that make up our modular world.

The later chapters have less to say to the general reader, although they are fresh and helpful to students familiar with the mathematics of wave scattering. The cartoon strips of time development are no longer available, since even at this level the physics demands two or three spatial dimensions. The plots therefore revert to steady-state representation, just as they do in the familiar time-free optics. Here the emphasis is on the spatial regions of physical interference that lie behind the common features of scattering cross sections, particularly resonance peaks. A little more attention to the time delays imposed on particles traversing interacting regions in one dimension might have made an easier bridge to the more demanding, if more realistic, considerations of resonances in the higher-dimensional cases.

Quantum states of well-defined energy do not vary in time. A good set of particle-probability clouds is plotted here for comparison with electrostatic reality. They depict both the real hydrogen atom and a couple of idealized model systems, in which simpler binding forces prevail. The book ends with a compendium of observed cross-sectional curves. They range from infrared absorption in paraffin molecules, through neutron scattering from lead nuclei, up to the peaks that mark the transient appearance of quark and antiquark pairs in the scattering of electrons from positrons at 10 gigavolts of kinetic energy. The phenomena are plainly kin over a dozen orders of magnitude in energy. The quantum edifice erected before 1930 remains as sturdy and beautiful as when it was new.


Astronomical Savings!
NIGHTWATCH
AN EQUINOX GUIDE TO VIEWING THE UNIVERSE
UNIVERSE GUIDE TO STARS AND PLANETS
PRACTICAL ASTRONOMY WITH YOUR CALCULATOR
PETER DUFFETT-SMITH
HOW TO USE AN ASTRONOMICAL TELESCOPE
JAMES MUIRDEN

Established 1845 SCIENTIFIC AMERICAN March 1986 Volume 254 Number 3

Rethinking Nuclear Power

A possible strategy for freeing nuclear power from its current impasse would be built around a new generation of lower-power, centrally fabricated nuclear reactors designed for inherent safety

by Richard K. Lester

Nuclear reactors now generate nearly 15 percent of the world's supply of electricity. Nations such as France, West Germany and Japan, already heavily reliant on nuclear power, are planning further nuclear growth. So are most Eastern European nations as well as some less developed countries. Yet in the U.S., where light-water reactors (LWR's), which provide most of the world's nuclear power, were first developed and where by far the largest number of LWR's have been built, nuclear power is at a dead standstill. U.S. electric utilities, which in the early 1970's were enthusiastically embarking on large numbers of nuclear projects, had largely abandoned the idea of building any more nuclear plants a decade later. Today, in spite of an apparent need for large amounts of new generating capacity by the end of the century, nuclear power cannot even claim to be, in former President Jimmy Carter's doleful phrase, the "option of last resort." To the nation's utilities it is not an option at all.

That there is nothing inherently unworkable about nuclear power is borne out by the success of nuclear projects overseas and indeed of many projects in the U.S. Moreover, although a variety of alternative sources of energy are available to utilities, and several promising new nonnuclear technologies are under development, abandoning nuclear power might well render the nation's electricity supply substantially less efficient and environmentally benign. Oil and gas, albeit currently plentiful, will eventually grow scarcer and costlier. Coal is difficult to burn cleanly: acid rain and other by-products of coal combustion are causing serious damage to the environment and will be expensive to control, and in the future the carbon dioxide released by the combustion of fossil fuels may have a severe effect on the global climate. The potential of solar electric technologies to compete economically outside a fairly narrow range of favorable sites or specialized uses has not yet been demonstrated.

Under these circumstances giving up on nuclear power would not be a wise course for the nation. Nuclear power in America faces problems that are so pervasive and severe, however, that it is not likely to be revitalized without fundamental changes in both industrial organization and the technology itself. The changes will be difficult to make. Paradoxically, however, the nuclear industry's current troubles and bleak prospects hold an opportunity: a chance to reexamine old assumptions and explore new initiatives. What steps could reverse the fortunes of nuclear power? And how likely is it that government and industry will take them?

The partial meltdown of a power reactor at Three Mile Island, Pa., in 1979 is commonly thought of as marking the turning point in the fortunes of the nuclear industry. The health effects of the accident were undetectable, but the fact that it took place at all and the often clumsy actions of the authorities in its aftermath dealt a sharp blow to public confidence in the safety of the technology. Moreover, by nearly destroying the financial viability of the utility that owns the plant the accident greatly heightened the sensitivity of electric utilities to the risks of nuclear investments.

For more fundamental reasons the enthusiasm of the utilities for nuclear power had begun to fade well before the accident. Modern nuclear plants are costly and large, with generating capacities of about 1,000 megawatts. The system of economic regulation to which utilities are subject has recently served to discourage investment in large capital-intensive power plants of any kind. For privately owned utilities, which supply almost 80 percent of the nation's electricity, state public-utility commissions and the Federal Energy Regulatory Commission control the return on investment by regulating rates. The regulators grant monopoly franchises to the utilities, which are obliged to provide an adequate supply of electricity to all consumers at reasonable cost; in return they are allowed to charge prices that bring their shareholders an adequate return.

For much of this century, when the price of electricity in real dollars was falling, the arrangement worked fairly well. Since the early 1970's, however, when the cost of power from new plants began to exceed the historical average price of electricity, pressure from consumers has increased and relations between utilities and regulators have soured. According to the utilities, revenues have often failed to keep pace with rising costs. More recently a growing number of state commissions have refused to allow utilities to recover in rate increases the full construction expenses for projects deemed in hindsight to have been unnecessary or poorly managed. (Many of the projects at issue are nuclear power plants.) These developments have deeply eroded investor confidence and have led most utilities to avoid major new construction of any kind.

Many of the other factors responsible for the misfortunes of the U.S. nuclear industry are unique to nuclear power. The continuing absence of facilities for the disposal of radioactive waste discourages utilities from placing new reactor orders, and a congressionally mandated timetable for the siting of both high- and low-level waste repositories continues to slip. Moreover, with some notable exceptions the performance of nuclear power plants has fallen well below expectations. Average annual plant-capacity factors (the percentage of their theoretical total output that nuclear plants actually produce in the course of a year) have for several years failed to exceed 60 percent, a figure well below the performance recorded in many other countries, including Japan, Sweden, Switzerland and West Germany.

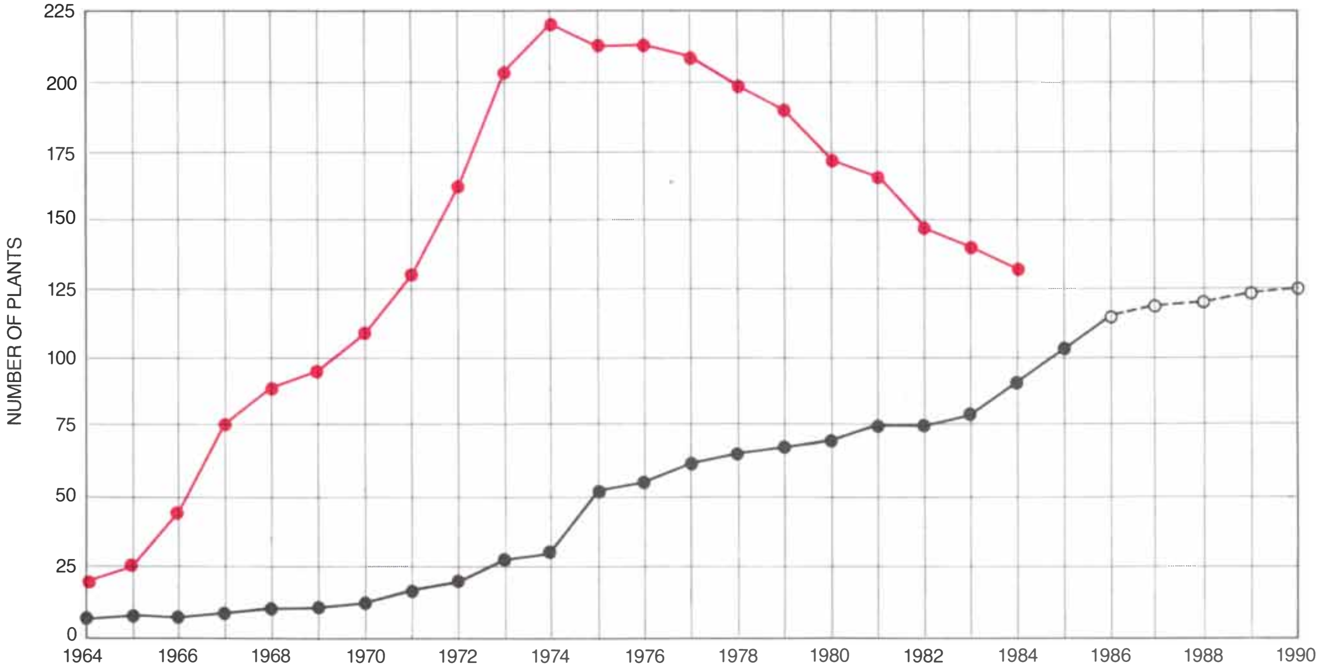
The most important reason utilities regard nuclear power as an unattractive prospect is that building a nuclear power plant is likely to cost too much. The 22 nuclear plants that began commercial operation in the period from 1983 through 1985 attest to the problem: their construction costs amounted to nearly $2,300 per kilowatt of generating capacity. In comparison, coal plants that entered service during a similar period cost on the average less than $1,000 per kilowatt. Such a large difference in capital costs between nuclear and coal plants more than offsets the lower fuel costs of nuclear power and has reversed its traditional economic advantage over coal.

There are exceptions to this picture, to be sure: several recently completed nuclear plants are cost-competitive with coal. It is typical cases rather than exceptions that shape the expectations of utilities, however, and the figures for new plants are becoming if anything still more discouraging. Another 10 plants are expected to enter service in 1986, and according to recent estimates their average cost will exceed $3,100 per kilowatt.

Why have nuclear plants become so costly? A rapid increase in construction time is one cause; the nuclear plants that entered service from 1983 through 1985 took an average of 10 years to build, in contrast to the five years required for the coal plants that began operation during a similar period and the five to six years needed for similar nuclear projects in France and Japan. The impact of long lead times has been compounded by the high inflation and interest rates of recent years. Even in today's more benign economic conditions, if two identical plants were started at the same time and one was built in five years whereas the other required the same amount of construction effort and materials but took 10 years to complete, the final cost of the second plant would be about 40 percent higher, solely because of inflation and increased financing costs. Many analysts have identified reducing construction time as the highest priority for the industry.

The influence of increases in the costs of physical inputs—materials, equipment and labor—on total plant cost is not as widely recognized. Yet these direct costs, in dollars per kilowatt of generating capacity, have grown over the past decade at an annual rate of 8 percent in real dollars, and even if it were not for the extended construction time, it is now questionable whether new nuclear plants in general could compete with coal-fired ones. Moreover, the only two U.S. nuclear units to be completed recently in less than seven years indicate that rapid construction is no guarantee of a low-cost plant: one of the two, the River Bend unit in Louisiana, is among the most expensive ever built. It is clear that more will be needed to restore the economic viability of nuclear projects than simply reducing construction lead times.

What has caused the rapid increase in direct costs? All input categories



CHANGING FORTUNES OF NUCLEAR POWER in the U.S. are evident in graphs of the number of nuclear plants on order (*color*) and in operation (*black*). Since the early 1970's, when orders for plants reached a peak, few utilities have called for new plants and many have canceled earlier orders. (The last plant order not subsequently canceled was placed in 1973.) Nevertheless, the number of plants in service has increased steadily. The figures for the number of plants operating in 1986 and beyond (*circles*) are estimates.

have contributed to the increase, but labor requirements have grown most spectacularly. According to recent estimates, the average amount of craft labor required for nuclear construction, measured in worker-hours per kilowatt of capacity, has increased fourfold since the early 1970's, and the needed investment of engineering and technical services has risen by a factor of more than 10. These two labor categories now account for two-thirds of the direct costs of nuclear construction. The growth of the labor requirement may have resulted in part from the general decline in the productivity of U.S. construction labor. In large measure, however, the trend reflects a potent combination of technological, organizational, managerial and regulatory factors that are unique to the nuclear industry.

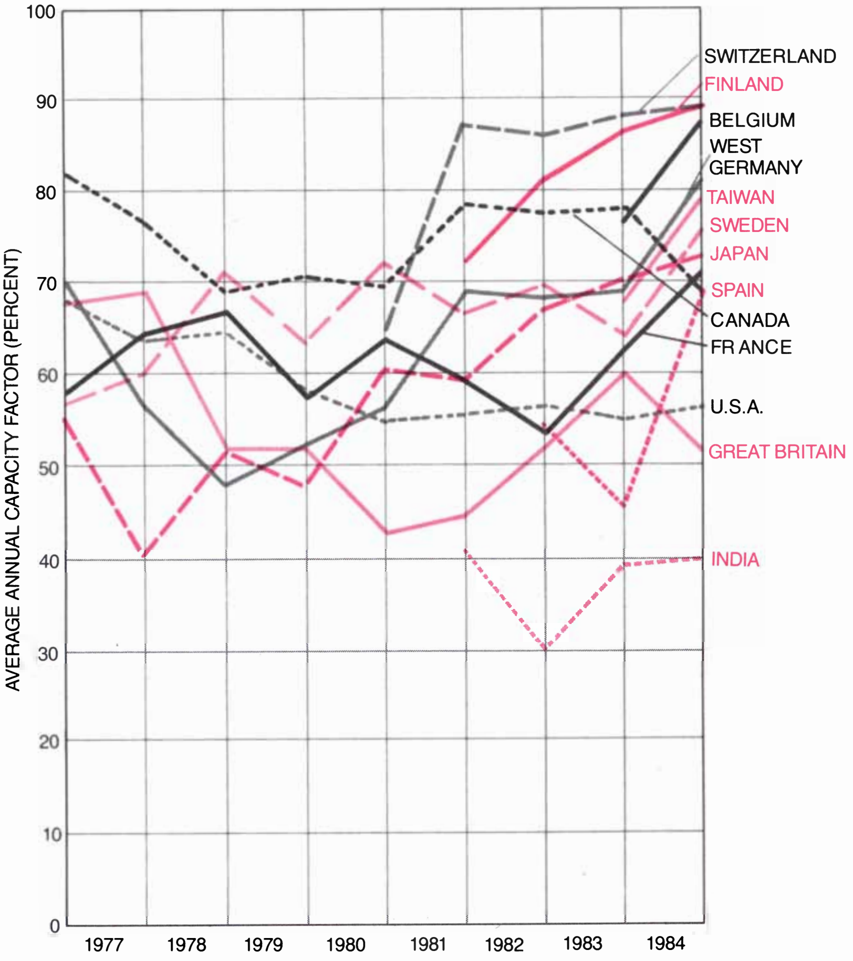
Nuclear plants have become much more complex in the past decade. Efforts to enhance reactor safety account for much of the complexity; a wave of new regulations promulgated since the early 1970's has affected virtually all aspects of plant design and construction. Not only have systems and components multiplied but also design and construction reviews have grown more elaborate and regulatory documentation has become correspondingly more voluminous. Moreover, the Nuclear Regulatory Commission (NRC) has applied many new requirements retroactively to plants under construction, making it necessary to redo much completed work.

Because of its disjointed structure, the U.S. nuclear industry has been hard pressed to cope with this increasingly complex technology. As in the construction of fossil-fuel plants, the task of designing and engineering a nuclear plant is divided among several organizations: a reactor vendor is responsible for the reactor and the other main nuclear systems and components, and an architect-engineer is responsible for most of the rest of the plant. A third contractor often undertakes the actual construction of the plant. Coordinating several independent organizations, each of which has different contractual terms and a limited technical purview, during a period of rapid technical and regulatory change would tax the skills of even the most seasoned project manager.

Because the electric-power industry is also disjointed, few utilities have been able to gain the necessary experience. The nuclear program of almost half of the 55 companies with nuclear units in operation or under construction consists of only a single plant. Moreover, a utility that builds a second plant may not profit fully from its earlier experience, because it will often be working with a different reactor vendor or architect-engineer. Four reactor vendors and 12 architect-engineers have taken part in the U.S. nuclear program, and several utilities have acted as their own architect-engineer. Only in rare cases has the same group of principals collaborated on several units in succession, a circumstance that has greatly retarded movement down the learning curve.

Variations in experience seem to explain much of the strikingly wide range in the direct capital costs of the most recent nuclear power stations: a variation of more than fourfold in cost per kilowatt. The least expensive plants were built by a utility, a vendor and an architect-engineer that had collaborated on several earlier projects. Most of the costliest units belong to utilities with no previous nuclear projects, and the remaining ones are the products of first-time collaborations between utilities and their suppliers.

The fragmented industrial structure has increased direct costs in another way: it has led to a lack of standardization in the technology. Indeed, there are almost as many power-plant designs as there are plants. The experience in France, where the national utility and a single supply consortium have followed a strict policy of building long series of very similar plants, attests to the benefits forgone in the U.S. Design and engineering costs are



AVERAGE ANNUAL CAPACITY FACTORS of nuclear power plants in various countries indicate the percentage of their theoretical total output actually produced by the reactors each year. The reasons for the comparatively low efficiency of nuclear power in the U.S. are not yet clear; in part it probably reflects the fragmentation of the nuclear industry, which has hampered the transmission of technical information that might have improved efficiency, and the lack of standardization in the technology, which limits the applicability of any given technical improvement. The utility-sponsored Institute for Nuclear Power Operations has begun to play an important role in fostering greater coordination, and some U.S. reactors have consistently achieved performance records higher than the national average, apparently benefiting from strong management of plant operations and maintenance.

lower when they are spread out over many power plants; engineers and construction workers grow more experienced and hence more productive, and suppliers can use their manufacturing capacity and training facilities more efficiently. Although few other national industries are as monolithic as the French program, none approaches the degree of disaggregation found in both individual construction projects and the industry as a whole in the U.S.

The fragmented industrial structure and the lack of standardization in the technology have also complicated the task of the NRC, which has confronted a large population of utilities with widely varying capabilities, operating plants with varying characteristics. The NRC, moreover, is beset with internal problems rooted in its unwieldy organizational structure. Indeed, several independent investigative committees, including the Kemeny commission, which examined the Three Mile Island accident, have concluded that the agency may be inherently ineffective in its present form. Its internal difficulties have contributed to arbitrary and inconsistent dealings with licensees, which have increased the cost and financial uncertainty of building nuclear plants. Industry officials have said repeatedly that until the financial risks associated with licensing new plants can be lowered, further nuclear orders are out of the question.

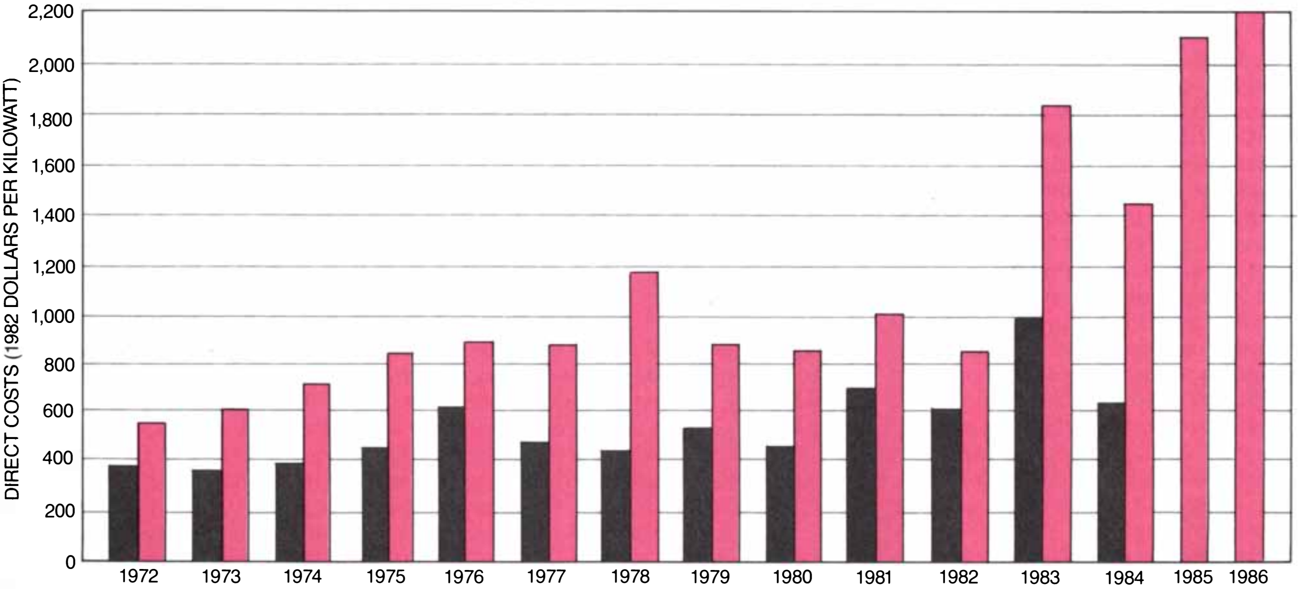
If a competent team of designers, well informed about utility capabilities and preferences, regulatory trends and public opinion, were today to set out from scratch to develop a nuclear plant system tailored to the U.S. electric-power market, would the result resemble current LWR's? The resemblance probably would not be close: the design would be likely to differ from the power reactors of today in three key respects.

First, the generating capacity of the new reactor would be considerably less than the 1,000-megawatt capacity characteristic of present-day nuclear plants. Utilities built large units in the hope of realizing economies of scale. The results have been inconclusive, and some recent studies suggest that the greater complexity of plants larger than about 800 megawatts may outweigh any such economies. Large plants have a clear economic disadvantage as well: by making utilities less likely to build several nuclear power plants in succession, they reduce the number of opportunities to gain and benefit from experience.

The present climate of economic regulation discourages investment in such large, capital-intensive plants, and most U.S. utilities are in any case too small, and are growing too slowly, to accommodate 1,000-megawatt units easily. Such plants make matching a utility's expansion in capacity to future growth in demand more difficult, particularly now that demand growth itself has become less predictable. In theory the joint ownership of large plants by several utilities could overcome these problems. As the participants in several such ventures have recently found, however, what may at first appear to be a way of spreading risks can turn into a mechanism for compounding them, particularly when the utilities are subject to different state regulatory commissions.

In the future, competition in the markets for wholesale power is likely to increase among established utilities and also between utilities and new power producers such as large industrial cogenerators, which produce both electricity and heat for industrial processes. Heightened commercial risks will generally serve to discourage orders for large central-station plants except by producers of proved competence and efficiency, and even they are likely to prefer adding capacity in increments of less than 1,000 megawatts.

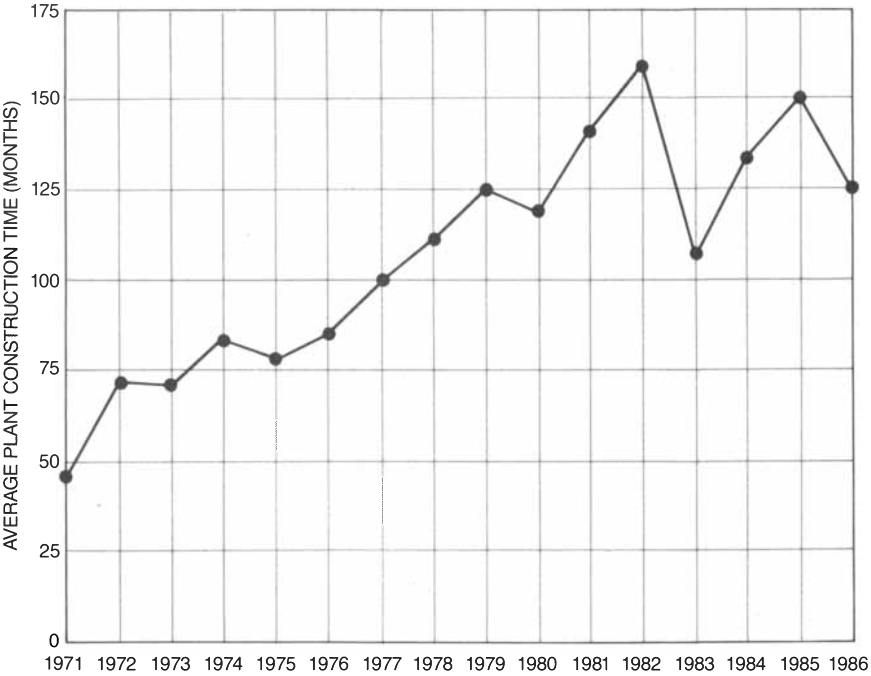
The second key difference between the hypothetical design and current practice would affect plant construction. The design would allow much of the fabrication to be completed in a central manufacturing facility instead of at the construction site, as is done now. The result would be an increase in worker productivity and work quality, which typically are higher in the controlled environment of a factory than they are in the field. Centralized production would make it possible to apply advanced automated manufacturing and inspection techniques.



DIRECT COSTS OF CONSTRUCTION for coal-fired (*black*) and nuclear power plants (*color*) that entered service each year from 1972 through 1984 are given in 1982 dollars per kilowatt of generating capacity. Projected direct costs are also shown for nuclear plants beginning operation in 1985 and 1986. Direct costs are a measure of the actual labor and materials required for construction; they do not include financing costs. Institutional problems and the enormous complexity of current nuclear power plants have combined to drive their direct costs well above those of coal-burning plants, more than offsetting the lower fuel costs of nuclear power.

In perhaps the most important of its advantages, such centralization would lead naturally to design standardization and would hasten the buildup of experience, with its associated benefits in efficiency and economy. Constraints on the weight and volume of components that must be transported suggest that centralized production might favor smaller power-plant size, but larger plants could also be designed to a greater degree than they are now for production as prefabricated modules, to be assembled later at the construction site.

The third novel feature of the design would be its safety philosophy. The goal of the change would be a reactor that was easier to site, more readily licensable and less susceptible to regulatory change than current LWR's are. The trouble with current reactors is not that they are unsafe; numerous technical assessments have shown that they pose much less of a threat to public health and safety than other commonly recognized hazards do. The difficulties arise because the technical basis of these findings is obscure and their accuracy is hard to demonstrate.



AVERAGE CONSTRUCTION TIME for U.S. nuclear power plants entering service each year from 1971 through 1986 has risen from less than five years to more than 10—about twice the time needed to build a coal-fired plant or to complete a nuclear plant in France or Japan. The extended construction time of nuclear projects has made them particularly susceptible to the effects of high interest rates and inflation and has driven up their total cost.

The safety of any reactor depends on its ability to prevent significant amounts of radioactivity from escaping to the environment. Because the radioactivity is normally contained in the fuel, the possibility of a major radioactive release arises only if core cooling fails and the fuel overheats or melts. Reactor cores are designed so that they cannot sustain fissioning once the temperature climbs much above its normal level; the reactor shuts down automatically when the core overheats. The inventory of radioactive fission products in the core continues to release large amounts of decay heat, however; if the heat is not removed, the temperature can continue to rise, leading to fuel damage. Even then additional barriers that seal off the core from the environment, including for example a stout containment building, may prevent the escape of radioactivity from the reactor.

Current LWR's derive their safety from a complex hierarchy of systems, some of them intended to prevent accidental overheating of the fuel and others to forestall the release of radioactivity in case cooling fails and the core is damaged. Some safety features, such as the containment building, are passive, but many of the systems, such as emergency sprays that force water into the core, must be activated, either automatically or by an operator. Because neither operators nor electromechanical components are completely reliable, designers have provided multiple backup systems to perform key safety functions. Multiple backups cannot completely eliminate the possibility of a major radioactive release, but they have reduced the probability to a very low level.

Calculating the residual risk is a complex matter, however, which requires identifying all the possible sequences of failures, electromechanical and human, that could lead to the release of radioactivity and estimating both their individual probabilities and their consequences. Such probabilistic risk assessments are unintelligible to most of the public. Moreover, major radioactive releases are so rare that it is not easy to verify the calculations by comparing them with actual experience. (The situation is different in other fields to which the same methods of risk assessment are applied, such as aircraft safety.) The perceived credibility of the investigator or agency conducting the evaluations thus becomes practically the only basis most people have for judging reactor safety. Inevitably, events that appear to undermine that credibility, such as the accident at Three Mile Island, assume a significance that far exceeds their technical impact.

In part because of the novelty and complexity of the techniques of probabilistic risk assessment, they have not had a central role in regulatory decision making until quite recently. Instead regulators have relied on engineering judgments to set detailed technical and procedural standards for individual components and have made compliance with all the requirements a proxy for overall safety. The lack of a firm quantitative foundation for judging the safety value of individual requirements has given the regulatory process an ad hoc quality that has troubled both the industry and its critics. Moreover, the reliance of the NRC on a large and increasing number of regulations setting out, in exhaustive detail, just how plants should be designed, built and run instead of simply specifying general safety standards has brought about an unhealthy transformation in the relation between the industry and the regulator, in which the users of nuclear technology have ceded much of the responsibility for its management to a remote regulatory bureaucracy.

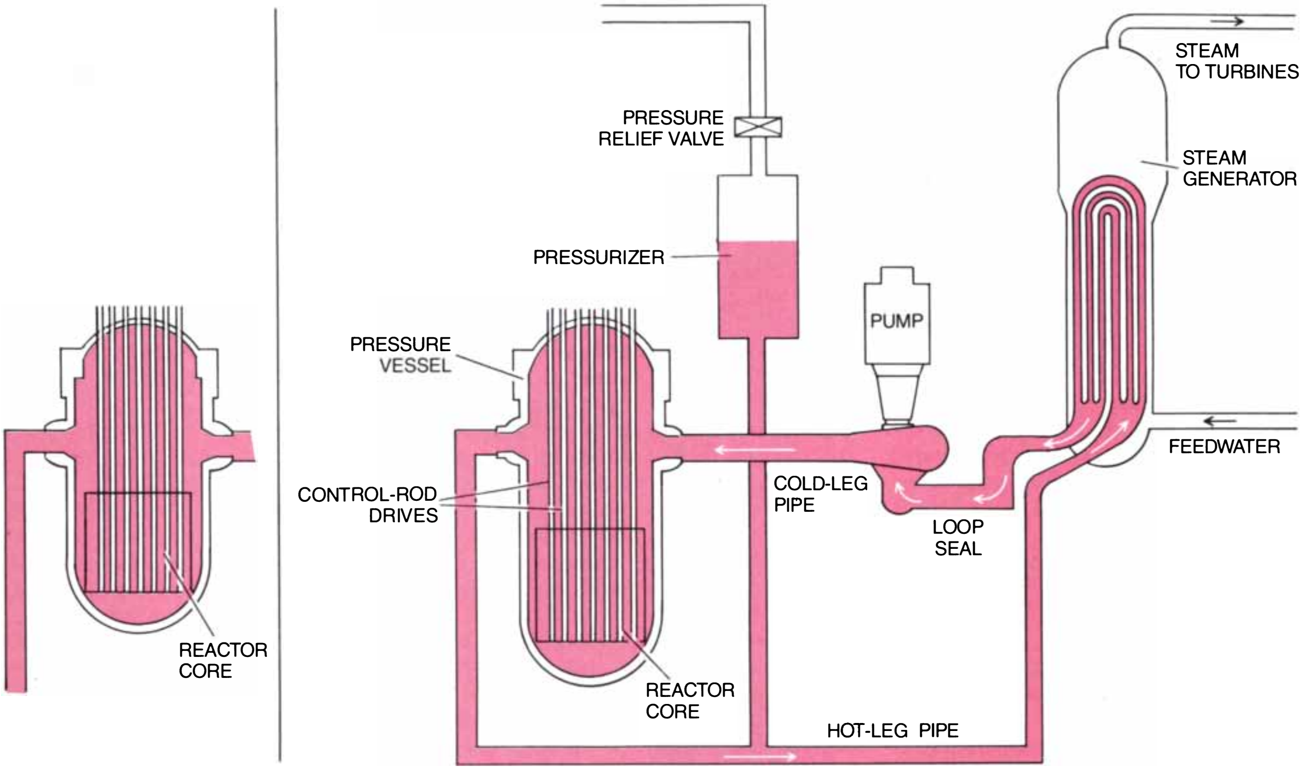
For the industry, the imperative of regulatory compliance has become paramount. Its creativity and initiative have been sapped. There is little room even for innovative approaches to safety, since it is usually so much easier to follow accepted procedures than to gain approval of new ones. Unless the dominance that regulation has assumed as a consequence of the current approach to safety can be reversed, it is hard to see how nuclear power can regain commercial viability.

What changes in reactor design could mitigate these problems? A promising approach that has recently attracted attention both in the U.S. and overseas is to design the reactor so that if normal cooling fails, the heat generated by fission products in the core can safely dissipate through natural heat-transfer processes such as convection or thermal radiation. The passive cooling would suffice to prevent overheating and core damage, eliminating the need for external forced-cooling systems and plant operators to activate them.

Such a strategy would considerably simplify the design and construction of the plant, and in particular its nonnuclear systems, which today are subject to strict regulatory standards because malfunctions there in theory can lead to core damage. A self-protecting reactor would make it possible for the rest of the plant to be designed and built to the standards of conventional fossil-fuel plants, and the savings in equipment and labor costs would be substantial.

Passive safety would also make judging safety risks a simpler matter, both by reducing the number of systems in a plant that need to be considered and by eliminating the troublesome problem of predicting the likelihood and effects of operator error. For the public the concepts of passive heat removal and "walkaway safety" would probably be more readily comprehensible than the complex safety features of conventional LWR's, and they could certainly be demonstrated more compellingly, by subjecting an unattended reactor to all kinds of simulated accidental and intentional disturbances. Finally, eliminating the risk of severe core damage could ease prospective plant owners' concerns about the enormous costs of coping with such an accident, even one with minimal health consequences.

Reactor designers have already begun to explore evolutionary changes to conventional LWR's in the directions of smaller size, increased factory fabrication and greater passive safety, and the utility-sponsored Electric Power Research Institute (EPRI) has undertaken a program to define technical requirements for the next generation of LWR's. The EPRI program, which focuses on large units, stresses the need for simplified plant designs and for reactors that would be easier to manage in an emergency. If coolant circulation is interrupted in a conventional pressurized-water reactor, the water in the core may boil away very quickly, leaving the fuel uncovered and allowing it to overheat. One of the steps toward greater passive safety that EPRI is considering is increasing the depth to

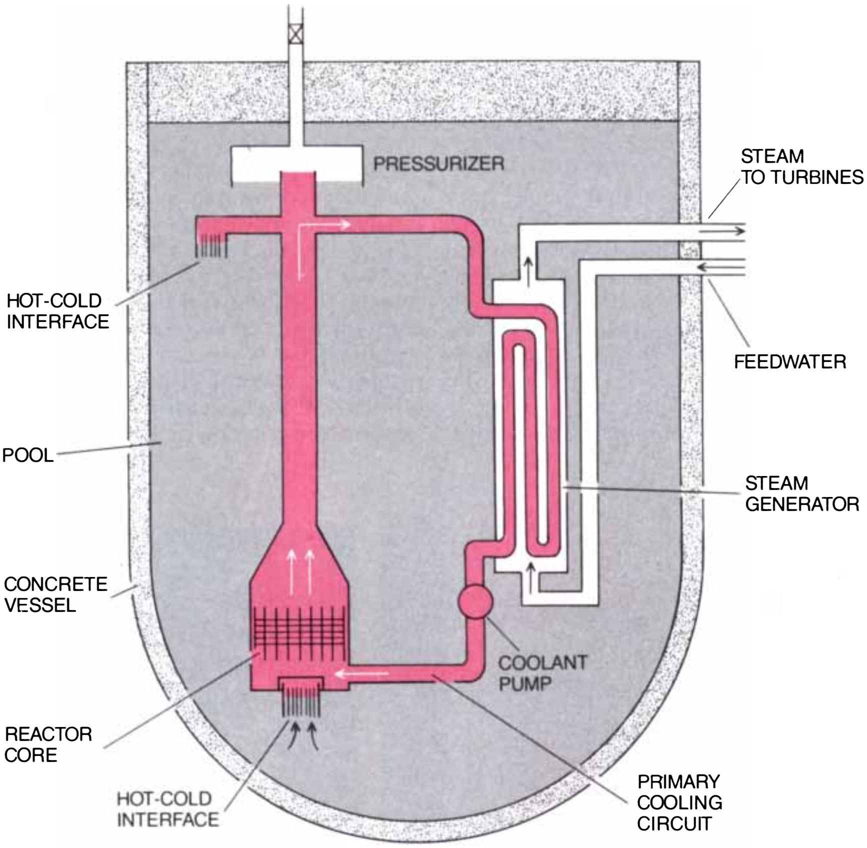


ADVANCED PRESSURIZED-WATER REACTOR, shown here in highly schematic form, shares its basic design with conventional pressurized-water reactors. Water in the primary cooling system flows through the cold-leg pipe into the pressure vessel and carries heat away from the reactor core through the hot-leg pipe to a steam generator, which sends steam to the plant's turbines. A pressurizer maintains the high pressure of the primary cooling circuit, preventing the coolant from boiling, and a loop seal protects the coolant pump from air bubbles. The new reactor design, the product of a collaboration among the Westinghouse Electric Corporation, the Mitsubishi Company, Japanese utilities and the Japanese government, specifies a larger core and a deeper pressure vessel than those of current reactors (*left*). The increase in core size reduces the power density of the reactor and consequently the density of fission products; thus the decay heat they release would raise the core temperature more slowly if cooling failed than would be the case in current reactors. The deeper vessel places the core well below the cold-leg pipe and the loop seal, ensuring that loss of coolant through a small rupture in the cold-leg pipe would not uncover the core—a possibility in the reactors in service today. In cooling failures of every kind the increased volume of water above the core would take longer to boil off than it would in present designs, extending the time operators would have during which to react to the emergency. The reactor, which would produce somewhat more power than most of today's reactors do, would embody many other improvements, meant to lower fuel costs, improve efficiency and ease maintenance.

which the core is submerged within the pressure vessel. In an emergency the water above the core would take longer to boil off than it would in current reactors, and damage to fuel elements would be delayed. Such measures would not eliminate the need for backup cooling and containment systems, but they would give operators more time to respond in an emergency than they have now.

New reactor concepts are emerging that go substantially further, particularly in the direction of passive safety. Among the most promising of these so-called inherently safe concepts is a small, high-temperature gas-cooled reactor (HTGR) developed mainly in West Germany. The reactor has a "pebble bed" core, in which the fuel is embedded in small graphite spheres. The fuel can withstand temperatures of up to 1,600 degrees Celsius without releasing fission products; the small size and large surface-to-volume ratio of the reactor ensure that the core temperature would not exceed that value even if the reactor lost all its helium coolant while operating at full power. Fissioning would stop automatically after an initial rise in temperature, as it would in a conventional LWR. Fission products would continue to produce heat, but passive heat loss from the walls of the reactor vessel would be sufficient to stabilize the temperature at a safe level. Because small size is crucial to the inherent safety of the design, such a reactor could produce no more than about 100 megawatts of electricity; a power station might consist of several HTGR modules. An HTGR of comparable size but different core design is under consideration in the U.S.

A Swedish concept, the Process Inherent Ultimately Safe (PIUS) reactor, envisions a radically reconfigured LWR. The reactor core, primary cooling system and steam generators are immersed in a large pool of cold, borated water within a prestressed concrete pressure vessel. The pool and the primary cooling system are hydraulically connected, but under normal conditions the pressure developed by the coolant pumps is just enough to keep the pool water from entering the core. Any disturbance in the cooling system would upset the balance, and the borate solution would flood the core. The boron in the water, an efficient absorber of neutrons, would shut down the chain reaction, and the cold pool water would carry off the residual heat. Neither an operator nor an electromechanical device would be needed to set these events in motion.



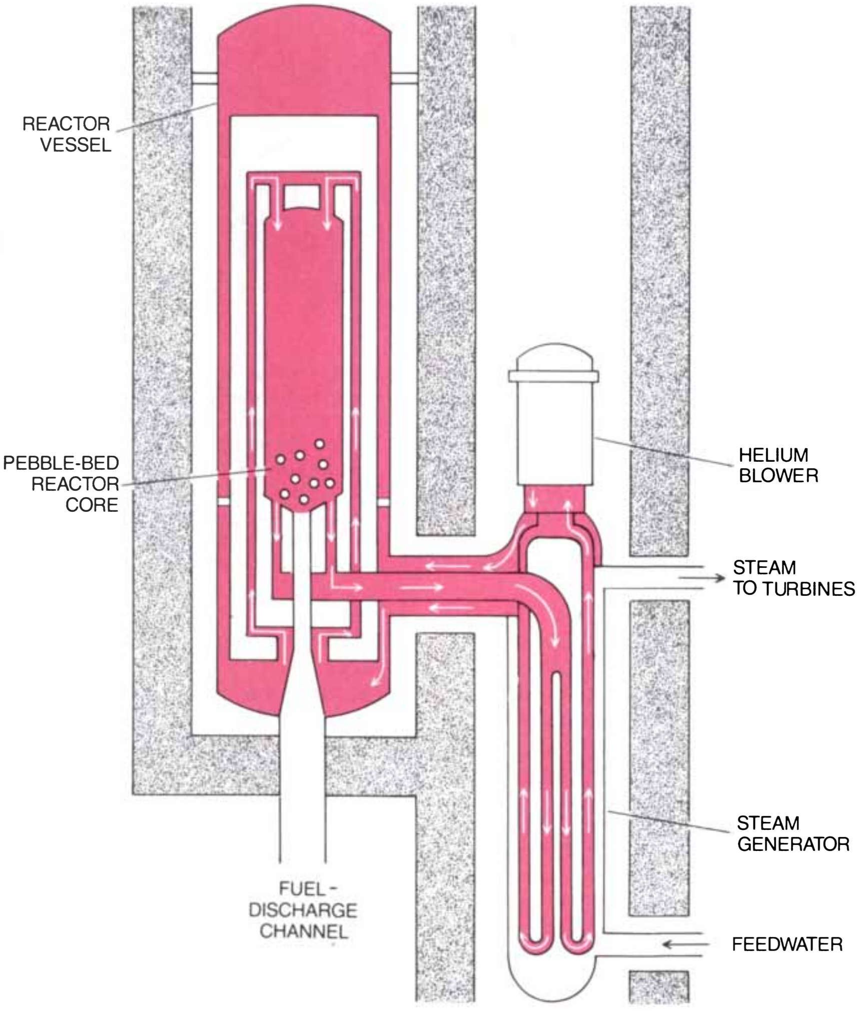
PIUS REACTOR, designed by ASEA-ATOM of Sweden, consists of a core and a primary cooling system similar to those of a conventional pressurized-water reactor but immersed in a pool of cold borated water. The pool and the cooling circuit are hydraulically connected at interfaces that are naturally stable because hot coolant lies above the denser pool water. In normal operation the pressure developed by the coolant pump keeps the pool water from flooding the core. If coolant circulation were to fail, natural convection would carry the pool water into the circuit (*black arrows*). The boron in the water would absorb neutrons, shutting down the chain reaction, and the large volume of water would suffice to cool the reactor for several days without the intervention of operators or emergency cooling systems.

A third kind of passively safe design has emerged from efforts to develop breeder reactors: reactors that produce more fuel than they consume by transforming isotopes that are unsuitable as nuclear fuel, such as uranium 238, into fissile isotopes such as plutonium 239. Such reactors, which use liquid sodium as a coolant, were conceived until recently in sizes ranging from 1,200 to 1,500 megawatts. The development of such large breeders continues overseas, particularly in France and the U.S.S.R. The expected steep increases in the price of uranium, which had been the main economic motivation for breeder development, now seem unlikely, however. U.S. designers have recently begun exploring much smaller liquid-metal reactors sized, like the modular HTGR, to enable them to dissipate heat passively in an emergency from the walls of the reactor vessel. Like large breeders, the smaller reactors would produce significant amounts of plutonium, which would be recovered by reprocessing the fuel.

The need for fuel reprocessing reduces the appeal of small liquid-metal reactors as a possible alternative to conventional LWR's. It would add significantly to the cost of commercializing the reactor technology and to the technical and organizational complexity of reactor-fuel production. Moreover, there is widespread concern that fuel reprocessing and the recycling of plutonium could contribute to the proliferation of nuclear weapons.

Many people in the American nuclear-power industry regard the idea of a major shift away from current technology as unrealistic. They argue that it is wiser to improve the technology incrementally, drawing on the rapidly growing store of experience with LWR's, than to turn to an unproved concept. Whatever the theoretical advantages of new systems are, they say, in practice the revolutionary concepts would be sure to encounter a long stream of unanticipated problems, just as LWR's did. The critics also question the wisdom of withdrawing from the international technological mainstream. France, West Germany, Japan and other countries are committed to LWR technology and have been devoting substantial resources to its improvement. By participating in international joint ventures, as two U.S. reactor vendors are now doing with Japanese partners, the U.S. industry can benefit from foreign technological advances even while the domestic industry remains depressed.

A more sensible strategy than adopting radical new technology, so the argument runs, would be to concentrate on dismantling the institutional barriers that prevent LWR's from achieving the same economic viability in this country that they have shown elsewhere. The aims of the strategy would include increasing the predictability of nuclear safety regulation and reforming the fragmented and disorganized industry. Greater standardization of reactor and power-plant designs and construction procedures would accompany the organizational and managerial steps. Some of the specific proposals call for extensive institutional reform: the creation of regional nuclear operating companies or even of a Federal nuclear authority. Others envision the consolidation of the nuclear-power-plant supply industry and the assumption by suppliers of more of the financial risk that building nuclear plants entails.



MODULAR HIGH-TEMPERATURE GAS-COOLED REACTOR is cooled by pressurized helium (*color*), which carries heat from the reactor core to a steam generator. In this design, proposed by KWU/INTERATOM of Germany, the fuel consists of tiny uranium particles, individually coated in layers of graphite and silicon carbide and embedded in "pebbles," or small graphite spheres. Hundreds of thousands of pebbles make up the core; new pebbles can be added continually while spent pebbles are discharged from the bottom of the reactor vessel, allowing the reactor to be refueled while it is operating. The fuel can withstand high temperatures without damage. Because the reactor is small and therefore has a large surface-to-volume ratio, heat loss from the outside of the reactor vessel by air cooling and thermal radiation would stabilize the core temperature at a safe level in the event of cooling-system failure. Such a reactor would produce about 100 megawatts of electricity; a full-size electric power station would combine several modules resembling the one shown.

The debate can be summarized in a fundamental strategic question: Which is more likely to be effective—an attempt to restructure political, industrial and regulatory institutions to accommodate the special demands of present technology, or an effort to tailor the technology to the capabilities, limitations and needs of the institutions that now exist?

The difficulty with the first approach is that the milder proposals seem unlikely to suffice, whereas the more radical schemes imply a centralized nuclear-power organization more akin to foreign programs than to the rest of the U.S. electric industry. The U.S. nuclear industry can go only so far in emulating its more successful foreign counterparts. The weakness of central planning and decision making in the U.S. economy, although it complicates the development of nuclear power, nonetheless reflects underlying social preferences that seem likely to persist.

The relative abundance of other domestic energy sources also sets the U.S. apart from foreign countries that are making a success of nuclear power. Thus nuclear power will continue to be seen in the U.S. not as the national economic and political imperative it has become elsewhere but as one of several competing energy technologies. A policy that seemed to give nuclear power preferential treatment or remove it from the competitive arena would not be likely to attract strong support. Any U.S. nuclear program must also reckon with the fact that the U.S. is a more open, litigious society than most; attempts to shield its industrial and regulatory institutions from direct public pressures are unlikely to work. Developing a new generation of products, better suited to U.S. market conditions, may in the end be a surer path to recovery for nuclear power.

A strategy of rethinking the technological options faces practical problems of its own. Even the alternatives to conventional LWR's that are well along in their development could not become commercially available until the next century. The high costs and risks of commercializing a new power-reactor technology would deter private industry from attempting to do so without significant Government support. Indeed, with most industry leaders still publicly committed to conventional LWR technology, the Government would inevitably find itself leading, and not merely supporting, an effort to develop new reactors. The massive budget deficits and the lack of public enthusiasm for nuclear power rule out such an effort for the moment.

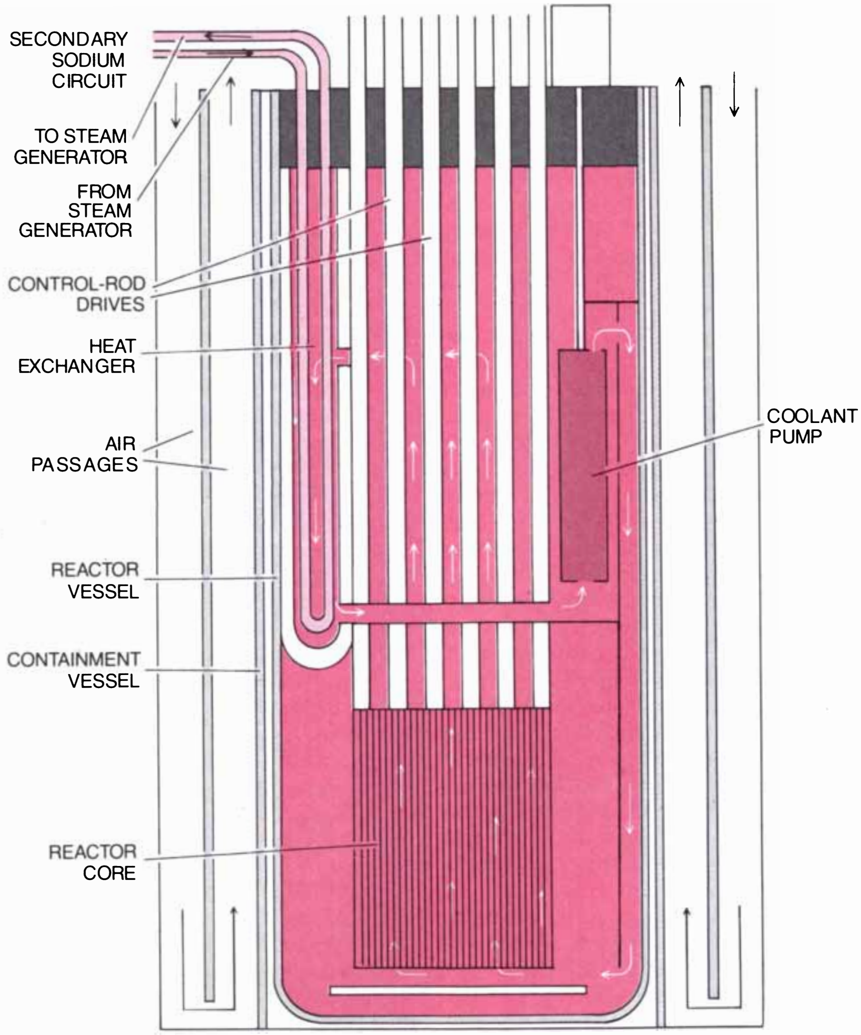
The attempt would be premature in any event. Much recent experience warns that in trying to choose a technology for commercial development the Government runs a considerable risk of developing something industry does not want and will not use. The willingness of the utilities themselves to make a substantial commitment to research is a prerequisite for a serious effort to develop a new generation of reactors. Such an effort will probably have to wait until regulatory reforms, institutional changes and incremental improvements in LWR's have been tried and the fate of current technology has become clearer.

In the meantime the Government and the nuclear industry should adopt the more modest goal of choosing, probably within the next three or four years, one of the promising passively safe technologies for development as a "proof of principle" device: a test plant intended not as a commercial demonstration but merely as a proof of technical feasibility. In particular, such a plant could demonstrate the principle of passive safety at a relatively modest cost. It could also serve as a vehicle for resolving key questions about the regulatory treatment to which passively safe systems would be subject. The engineering experience gained in the project would inform the later decision about whether to proceed to a commercial demonstration plant, and undertaking a proof-of-principle project now would allow that decision to be made in the mid-1990's, when the future of conventional LWR's will be considerably clearer than it is today.

The cost of such a project could be shared by the Government and a consortium of utilities and suppliers, including interested foreign suppliers; the willingness of suppliers and utilities to share in the development costs would aid in determining which of the candidate design concepts should be selected. The project could be done with total annual expenditures that were less than recent annual outlays by the Federal Government alone for breeder-reactor development.

The range of conceivable futures for nuclear power in the U.S. thus comes down to three broad possibilities. One is a return to LWR technology in an improved form, perhaps in as little as a decade. The revival of the technology would take place within a more streamlined industry, with fewer and more competent organizations operating in a stabler regulatory climate. Foreign LWR suppliers, having enjoyed livelier domestic markets in the meantime, would introduce many of the improvements and might figure commercially in the U.S. revival. In the second scenario conventional LWR's would fail to regain commercial acceptance, but after some years one or more second-generation reactor technologies conceived for small size, passive safety and centralized, modular fabrication would reestablish nuclear power as a major source of electricity for the next century.

At the moment it is not possible to say which of the first two scenarios represents the likelier outcome, and indeed there is no need to do so. Efforts to reform existing institutions and to improve conventional technologies should go forward along with explorations of radical new technologies. Neglecting either approach will make all the likelier the third outcome: the disappearance of nuclear power as an option for the future. That is a loss the nation can ill afford.



MODULAR LIQUID-METAL REACTOR has multiple cooling circuits. The liquid sodium of the primary cooling system (*dark color*) is forced through the core by a pump in the vessel; at heat exchangers the sodium gives up heat to secondary sodium circuits (*light color*), which drive steam generators (not shown). If the coolant pumps were to fail, sodium would continue to circulate through the reactor core because of natural convection. The reactor's small size would enable it to lose enough heat to air convecting through passages around the outside of the containment vessel to avoid core damage and the release of radioactivity. The General Electric Company proposed this design; it would yield about 140 megawatts of electricity. As in large liquid-metal breeder reactors, substantial amounts of nuclear fuel would be made in its core as neutrons transformed uranium 238 into plutonium.

The Earth's Magnetotail

The solar wind sweeps the earth's magnetic field into a vast tail. Disruptions of the tail generate bright auroras at the earth and propel great bodies of magnetized gas into interplanetary space

by Edward W. Hones, Jr.

The earth resides in a vast magnetic cavity. The cavity, called the earth's magnetosphere, exists because the solar wind, a gas of charged subatomic particles flowing continually from the sun, cannot easily penetrate the earth's magnetic field. Instead the solar wind stretches the field into a more or less cylindrical region extending from the earth into interplanetary space like a great wind sock millions of kilometers in length. This cylindrical region is the earth's magnetotail.

Our current knowledge about the magnetotail is the outgrowth of a long effort to understand the auroral lights that glow in the skies at high latitudes, an effort that enlisted such notable investigators as Galileo, Halley, Celsius and Franklin. Progress was relatively slow until the arrival of the space age, about 30 years ago, when satellites bearing scientific instruments began to aid the study of the interplanetary space surrounding the earth. This led to the discovery and exploration of the earth's magnetosphere. The explorations in turn have led to a broad conclusion: The solar wind and the magnetosphere form a vast electrical generator—one in which the interaction of magnetic fields and plasma (the gas of solar-wind particles) converts the kinetic energy of the solar wind's motion into electricity.

The electric power creates a wealth of phenomena. These include not only the beautiful and intriguing displays of auroras but also the Van Allen radiation belts, which surround the earth. Another of these phenomena, discovered in the magnetotail only three years ago from satellite data, consists of huge plasma structures called plasmoids. A plasmoid is a body of hot plasma threaded and held together by loops of magnetic field. Magnetic field lines surrounding the plasmoids propel them from the magnetotail, like projectiles from a cannon, at speeds of millions of kilometers per hour.

The interaction of the solar wind with the magnetosphere begins at the surface of the magnetosphere, which is called the magnetopause. Its sunward point, where the solar wind's forward momentum is stopped by the earth's magnetic field, lies 10 earth radii "upstream" from the earth. (The earth's radius, 6,370 kilometers, is commonly used as a unit of distance in magnetospheric research.) Here, in the region sunward of the earth, the earth's magnetic field is compressed by the interaction of the wind and the field. At the night side of the earth the field is stretched far downstream to form the magnetotail. The diameter of the tail is between 40 and 60 earth radii; its length exceeds 1,000 earth radii.

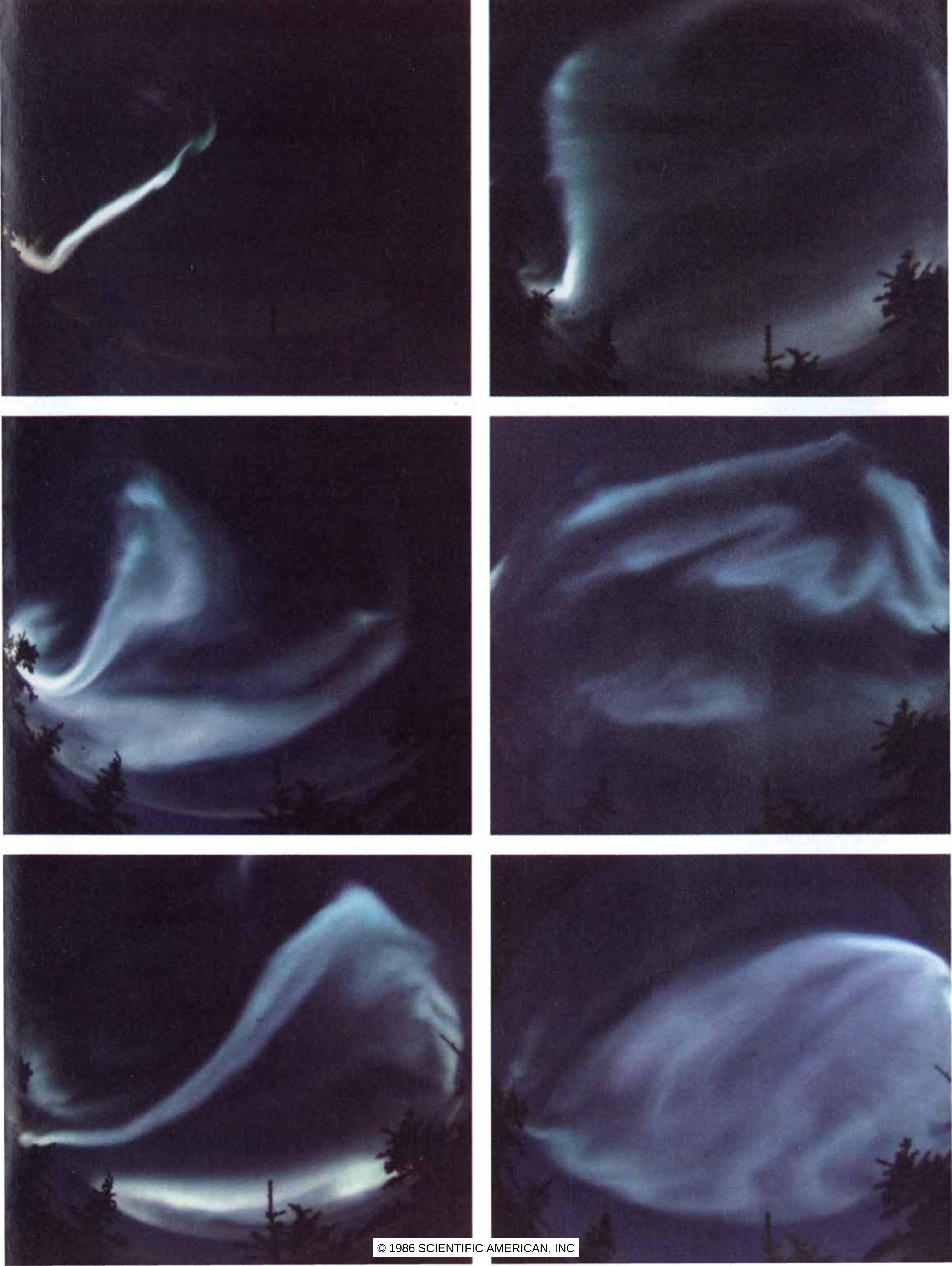
The magnetotail consists of adjacent halves called lobes, which have opposite magnetizations [*see illustration on pages 42 and 43*]. In the upper, or north, lobe the magnetic field points sunward and the field lines connect to the north polar region of the earth. (The direction of the field is the direction in which the north end of a compass needle would point if the compass were placed in the field.) In the lower, or south, lobe the field points antisunward and the field lines connect to the earth's south polar region. The two lobes are separated by a sheet of electric current that flows across the midplane of the tail and then loops around the north and south lobes. These loops of current create the magnetic fields in the lobes.

The solar-wind plasma is not completely excluded from this complex magnetic domain. Some plasma penetrates the magnetosphere's sunward region and populates the surface regions of the tail. From these regions the plasma flows through the lobes toward the midplane of the tail, where it forms a concentration of plasma called the plasma sheet. The plasma sheet is the site of the current separating the lobes.

The processes that sustain the magnetotail are interactions of charged particles, electric currents, electric fields and magnetic fields. Acting together in the magnetotail, they create a highly complicated physics. Taken by itself, however, each type of interaction is fairly straightforward. Thus an understanding of the individual interactions can give insight into the structure their interplay creates.

A magnetic field acts on a charged particle through what is called the Lorentz force. The force is proportional to the magnetic field strength and to the component of the particle's velocity perpendicular to the field and is directed at right angles to both [*see top illustration on page 44*]. In a uniform magnetic field the Lorentz force causes electrons and protons to move in circles, but in opposite directions. If a particle's original motion has a component parallel or antiparallel to the field's direction, the addition of circular motion will shape the particle's

ENERGY FROM THE MAGNETOTAIL released in the ionosphere, about 100 kilometers above the surface of the earth, is responsible for the aurora borealis, or northern lights, shown (*top to bottom, left to right*) in a sequence of photographs that Robert H. Eather of Boston College made at Churchill, on Hudson Bay in northern Manitoba, near midnight on an evening in March. Eather fitted his camera with a fish-eye lens, so that each image shows the entire sky from horizon to horizon; in each image the southern horizon is at the bottom. Electrons bombarding atoms in the ionosphere produce the auroral light. The electrons, expelled from the sun in a wind of charged particles (the solar wind), get trapped in the magnetotail and approach the earth when a "magnetospheric substorm" disrupts the magnetotail's resting configuration. The images span a 20-minute interval during such a disruption.

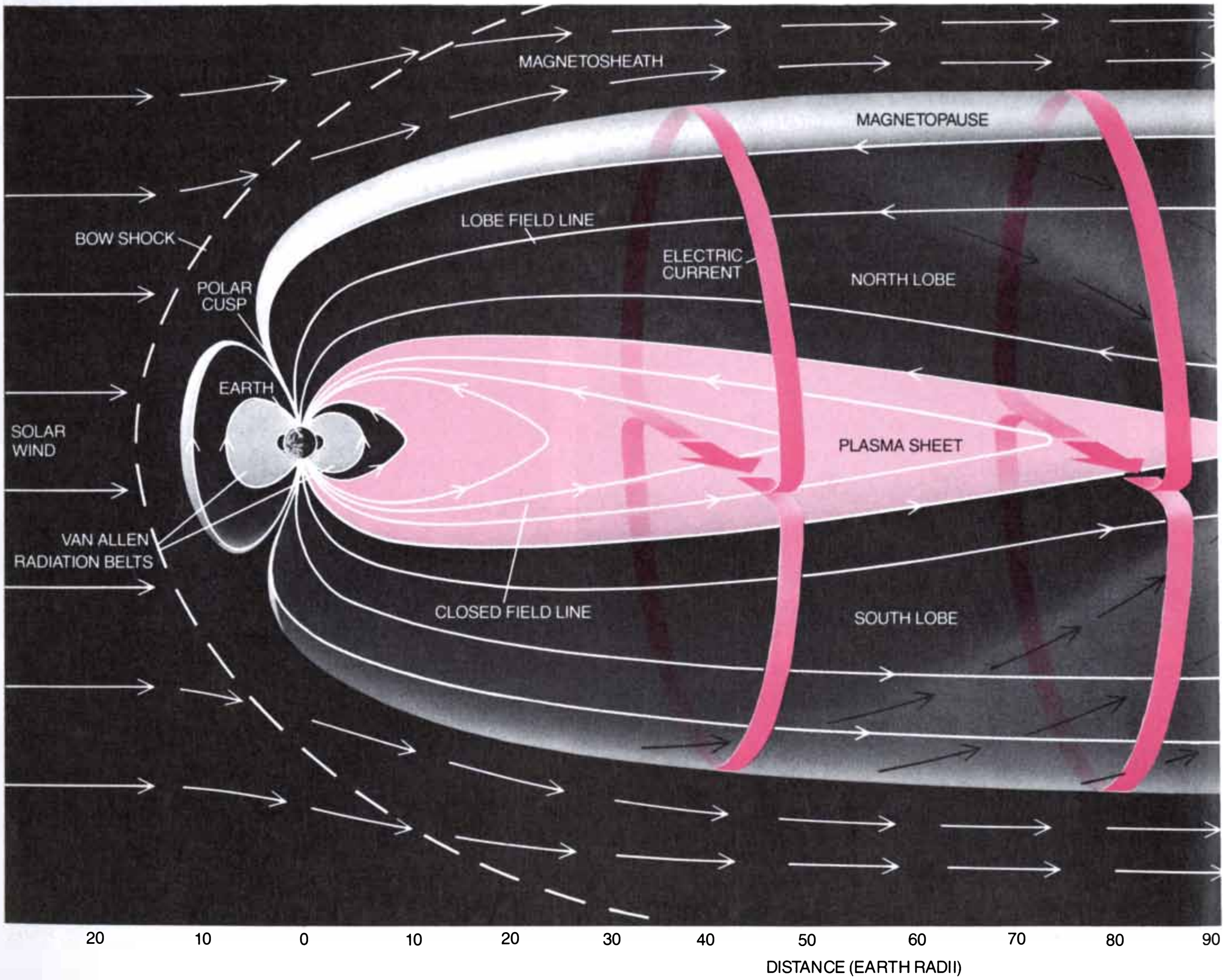


trajectory into a helix. As a particle moves, the center of its helical path is said to trace out a line of the magnetic field. In effect charged particles are "tied" to magnetic field lines: each particle spirals around a line.

A second interaction of charged particles and fields is called E-cross-B drift. When an electric field is imposed perpendicular to a magnetic field [*see middle illustration on page 44*], the spiraling particles tied to magnetic field lines experience a drift perpendicular to both the electric field (E) and the magnetic field (B). Positive particles such as protons and negative particles such as electrons drift in the same direction and at the same speed; hence the particles in a plasma (a gas of equal numbers of positive and negative particles) all drift as an ensemble. The magnetic field lines can be thought of as moving with the plasma; their speed is the E-cross-B drift speed of the particles. Under these conditions the magnetic field lines are said to be frozen into the plasma.

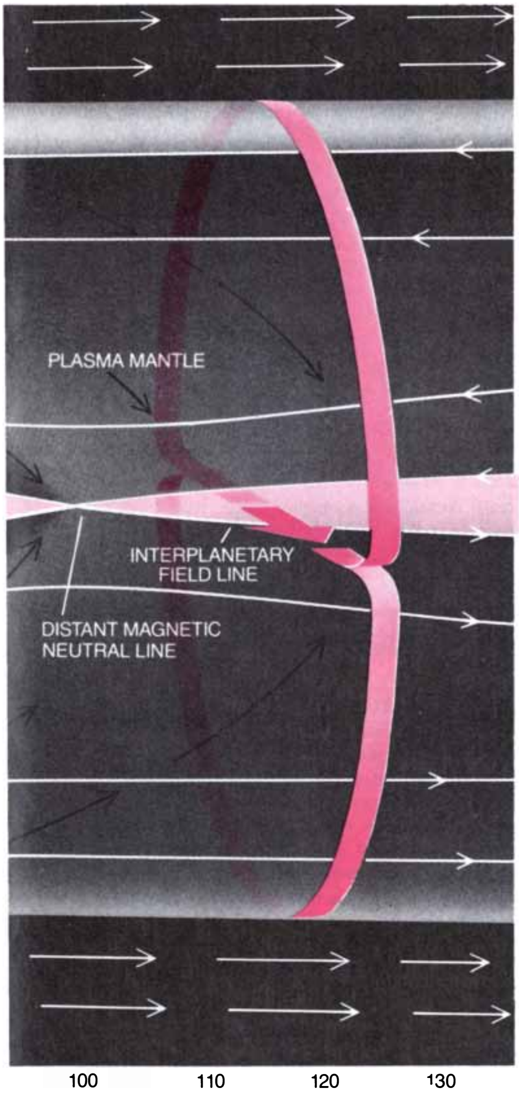
There are circumstances in which the magnetic field becomes very weak or changes very sharply in space or time; then the plasma particles and the field lines no longer move in tandem. Instead the field lines "diffuse" through the plasma. Under such conditions field lines of opposite direction can come together, break and reconnect in new combinations. The process, called magnetic reconnection, is crucial in the magnetosphere both for its acquisition of energy from the solar wind (a more or less continual process) and for its releases of energy from the magnetotail, creating auroras and plasmoids.

Magnetic reconnection takes place when regions of opposed magnetic fields come together. It can be visualized most easily in terms of the magnetic field changes that occur when two horseshoe magnets identical in size and strength are brought together so that their opposite poles align (that is, the north pole of one horse-



SCHEMATIC CROSS SECTION of the magnetosphere displays its chief magnetic and electric features. The solar wind is deflected along a front called the bow shock. It then flows around the magnetosphere in a region called the magnetosheath. The magnetotail is the part of the magnetosphere "downstream" from the earth; its surface is called the magnetopause. In its upper, or north, lobe the magnetic field lines point toward the sun; in the lower, or south, lobe the field lines point antisunward. Between the lobes a sheet of plasma (a gas of positive and negative particles) extends across the tail; it carries a cross-tail electric current (*color*) that loops around the lobes. A magnetic neutral line resides about 100 earth radii downstream from the earth; it too crosses the tail (the drawing shows only its midpoint). At the neutral line the lobe field lines, carried toward the plasma sheet by electromagnetic forces, meet and reconnect; the reconnection converts the lobe lines into closed field lines (loops tied to the earth) and interplanetary field lines, shoe faces the south pole of the other). The magnetic field arising from the magnets changes, so that some of the field lines that had originally run from the north pole to the south pole of one magnet now run from the north pole of one magnet to the south pole of the other.

Magnetic reconnection in the earth's magnetosphere is more complex and less amenable to intuitive insight. The complexity arises because the magnetic regions of the magnetosphere are plasmas, which are fluid and readily conduct electric currents. The currents create magnetic fields, which in turn reshape the plasmas. Nevertheless, the basic process is the same: magnetic reconnection takes place in the earth's magnetosphere when oppositely magnetized regions of plasma come together.



which extend out of the tail so that both ends reach into the solar wind. Magnetosheath plasma that enters the polar cusps occupies regions of the lobes called the plasma mantle. The plasma drifts toward the plasma sheet and is injected into the sheet by the reconnection process. The tail extends at least 1,000 earth radii into space.

One site of magnetic reconnection in the magnetosphere is the plasma sheet. In fact reconnection continually renews the plasma sheet. The process begins as solar-wind plasma gains entrance to the magnetosphere at sites such as the polar cusps (the places where the magnetosphere's magnetic field lines dip toward the earth over the north and south magnetic poles). Initially it flows downstream in the magnetotail along magnetic field lines in the outer levels of the north and south lobes, a region called the plasma mantle. As it flows it is subjected to an E-cross-B drift, which forces it toward the plasma sheet, at the midplane of the tail. The flowing plasmas carry with them the opposing, frozen-in magnetic fields. At some location along the interface between the lobes the frozen-in field condition breaks down, producing a diffusion region: a place where the field lines can move through the plasma.

Within the diffusion region, field lines of opposite direction begin to touch and reconnect across the plasma sheet. (Lines pointing sunward arrive from the north lobe and lines pointing antisunward arrive from the south lobe.) At the place where the opposing field lines touch, the net magnetic field strength is zero, owing to the presence of the opposite magnetizations. Hence the site of reconnection is called the magnetic neutral line. Normally it is situated in the plasma sheet about 100 earth radii downstream from the earth. On each side of the neutral line, plasma is ejected from the diffusion region along the midplane of the magnetotail in a narrow, wedge-shaped jet. The jet ejected earthward renews the plasma sheet; the jet ejected tailward flows downstream, back into the solar wind.

As a result of reconnection the magnetotail has three types of magnetic field line. (A fourth type, described below, exists only transiently and forms the magnetic structure of a plasmoid.) Each type occupies a particular region of the tail, and each type is characterized by its relation to the earth. The first type consists of the field lines in the lobes. One end of each of these lines is attached to the earth; the other end extends downstream into the solar wind. Such lines are called open field lines.

The second type is found in the plasma sheet on the earthward side of the magnetic neutral line. Here each field line comprises the earthward ends of two lobe field lines that have reconnected. Thus each field line in the plasma sheet on the earthward side of the magnetic neutral line is a loop, both ends of which are attached to the earth. Such lines are called closed field lines.

The third type is found in the plasma sheet downstream from the magnetic neutral line. Here each field line comprises the downstream ends of two of the lobe field lines that have reconnected; thus they are loops completely free of the earth. Their ends extend antisunward into interplanetary space. They are known as interplanetary field lines.

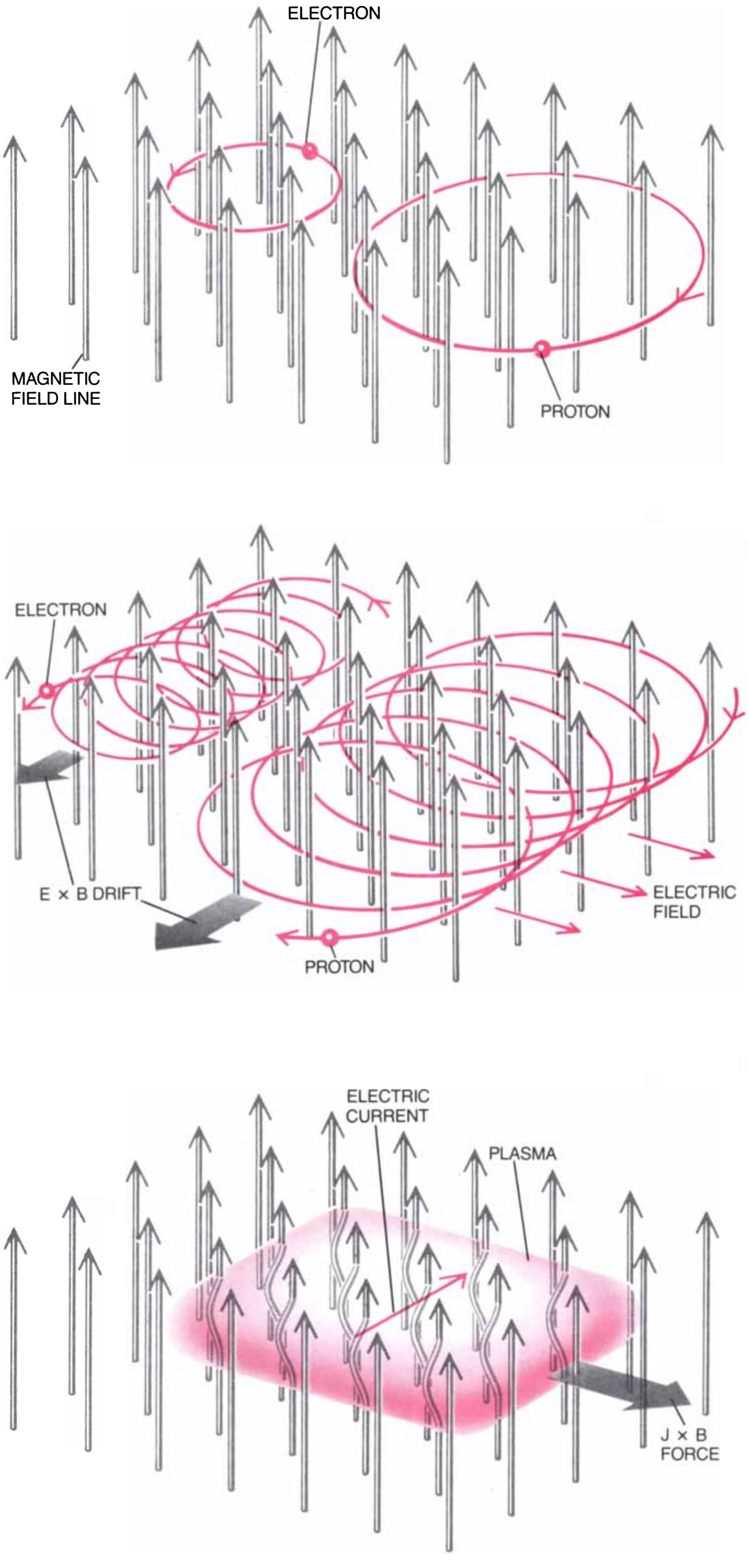
This is not to say the magnetotail always maintains a magnetic pattern consisting simply of these three types of magnetic field line. There are times when the interaction of the solar wind and the earth's magnetic field overloads the magnetotail with energy, leading to the phenomenon called a magnetospheric substorm, which disrupts the magnetic pattern. The substorm is the mechanism by which the magnetosphere intermittently releases large amounts of energy that has been stored in the magnetotail. Some of this energy goes to create the auroras, near the earth, while the rest is released downstream to the solar wind in the form of a plasmoid. (A substorm, which lasts for an hour or so, is distinguished from a geomagnetic storm, which lasts for a day or more and is caused when a solar flare initiates a shock in the solar wind.)

To understand the substorm, the aurora and the plasmoids one must understand the details of how the solar wind transfers energy to the earth's magnetic field and how the energy stretches, and finally overstretches, the field. The crucial point is that the encounter of the solar wind with the earth's magnetic field generates electricity because the solar wind and the magnetosphere constitute a magnetohydrodynamic, or MHD, generator.

In an MHD generator built by human hands an electromagnet generates a strong magnetic field between two plates, while plasma flows between the plates in a direction perpendicular to that of the field [*see illustration on page 45*]. The device that draws power from the generator is in an external circuit connecting the plates. As particles enter the field the Lorentz force deflects them, so that electrons move toward one of the plates and protons toward the other, creating within the plasma a migration of electric charge called the polarizing current. If the particles had enough space, the deflection would draw them into circular paths. Here, however, the charge-conducting plates intervene. One plate collects negative charge while the other collects positive charge. The charge leaves the generator and flows through the external circuit as a so-called depolarizing current.

In the magnetosphere the magnetohydrodynamic process generates electricity at two different locations. The more important generator arises as a consequence of magnetic reconnection at the sunward limit of the magnetopause. There the solar wind encounters the earth's magnetic field head on. The wind itself includes a magnetic field, which consists of field lines pulled outward from the sun. The direction in which the field lines point changes from time to time, seemingly at random. When the field lines point southward, the direction opposite to that of the earth's field lines at the sunward magnetopause, they can reconnect with the earth's field and so become tied to the earth [*see illustration on page 46*]. The solar-wind plasma, flowing through this earth-tied field, then completes the basic elements of an MHD generator. The earth's ionosphere (the ionized, electricity-conducting layer of the atmosphere about 100 kilometers above the earth) constitutes the low-resistance external circuit through which the polarization charge from the solar wind can flow.

When the solar wind's field points northward, as it does about half of the time, magnetic reconnection at the sunward magnetopause is much reduced and the flow of electrical energy diminishes. Even then, however, the second, less powerful generator persists in its activity. At low latitudes some solar-wind plasma flows across the closed lines of the earth's magnetic

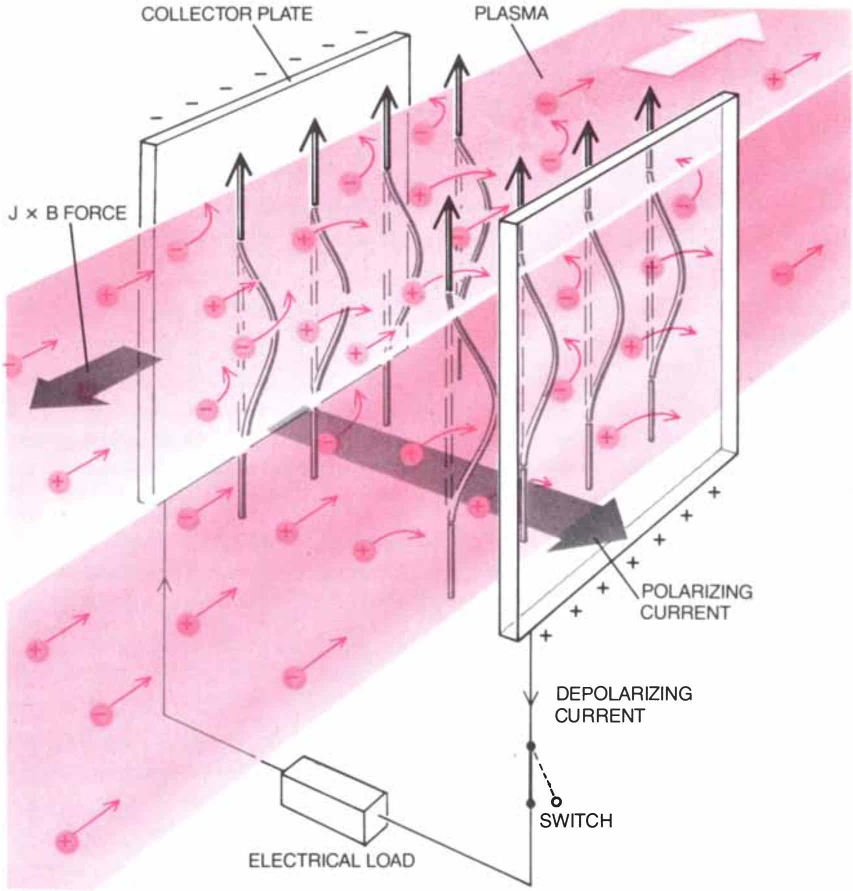


THREE FORCES that share in creating and sustaining the structure of the magnetotail are diagrammed. In a uniform magnetic field the Lorentz force (*top*) causes electrons and protons to move in circles in opposite directions, thus "tying" the particles to magnetic field lines. If an electric field is imposed perpendicular to the magnetic field, the charged particles acquire a further motion called E-cross-B drift (*middle*). The drift carries the centers of the particles' circular paths in a direction perpendicular to the directions of the fields. In this way plasma in the lobes of the magnetotail is driven toward the plasma sheet. Finally, plasma carrying an electric current that flows perpendicular to a magnetic field experiences a J-cross-B force (*bottom*). The force accelerates the plasma in a direction perpendicular to both the direction of the current and that of the magnetic field. The current is itself the source of a magnetic field, which distorts the original field, bending the field lines in the direction opposite to that of the force. The J-cross-B force resisting the flow of solar-wind plasma distorts the earth's magnetic field, producing the magnetotail.

field. The means of entry is not known, but it does not seem to depend strongly on the direction of the solar wind's magnetic field. In any case, the entry again creates the conditions for magnetohydrodynamic power generation: plasma flowing at a right angle to a magnetic field, with a polarization discharge path through the ionosphere.

Through these instances of magnetohydrodynamic generation the solar wind injects electricity into the earth's magnetic field. In turn the electric currents stretch the earth's field lines. The stretching results from an electromagnetic phenomenon called the J-cross-B force [*see bottom illustration on opposite page*]. Electric current (in this case the current produced by the magnetohydrodynamic process) flowing through a plasma at a right angle to a magnetic field exerts a force on the plasma in a direction perpendicular to the current (J) and to the magnetic field (B). At the same time it bends the magnetic field lines in a direction opposite to the direction of the force.

The degree to which the field lines are bent in an MHD generator depends on the strength of the magnetic field, the strength of the polarizing current and the size of the generator itself. In MHD generators being developed on the earth for electric-power generation the bending is inconsiderable. In the magnetosphere the bending is prodigious. If the solar wind were to stop, so that energy transfer no longer took place, the earth's magnetic field would simply assume the familiar symmetrical pattern of a dipolar field: the field of a bar magnet. Instead—and because of the J-cross-B force exerted by the wind—the antisunward side of the field becomes the magnetotail, a far more extensive magnetic arrangement and one that represents the storage of a great quantity of energy: about 10^{11} megajoules, the amount of electrical energy consumed by the U.S. in several days. The total rate at which the two magnetohydrodynamic mechanisms supply energy to the magnetosphere is between 10^5 and 10^6 megawatts, or roughly the rate at which the U.S. consumes electrical energy.



MAGNETOHYDRODYNAMIC GENERATION of electrical energy from the kinetic energy of a flowing plasma is the mechanism by which the solar wind transfers energy to the magnetotail. In a magnetohydrodynamic generator, plasma (*color*) encounters a magnetic field; the Lorentz force propels the plasma particles in opposite directions (depending on charge), toward plates at the sides of the field. The migration constitutes an electric current (*polarizing current*) perpendicular to the magnetic field; when the switch in the external circuit (*bottom*) is closed, the current flows continually. The J-cross-B force resists the plasma flow; the magnetic field caused by the polarizing current bends the device's field.

The deposit of some of this energy in the inner magnetosphere, near the earth, brings on intense auroral displays; it also injects energetic electrons and ions into orbits between about one and six earth radii from the earth, thus populating the Van Allen radiation belts. Still, a large part of the energy (and plasma) stored in the magnetotail is ultimately returned to the solar wind in the form of plasmoids.

Photographs encompassing the entire sky, made at one-minute intervals by a network of Arctic observatories during the International Geophysical Year (1957–58), have done much to establish the terrestrial side of this interconnected series of events. According to Syun-Ichi Akasofu of the University of Alaska at Fairbanks, who derived his model from the photographs, an auroral substorm (the aspect of a magnetospheric substorm that is visible in the polar regions of the earth) begins with the sudden brightening of an auroral arc. The auroras then spread eastward, westward and poleward, reaching magnetic latitudes of between 75 and 80 degrees north and south in about half an hour. Next comes the recovery phase of the substorm, which lasts for an hour or so, during which the auroras relax Equatorward. Observations made concurrently in the north and south polar regions show that the substorm evolves simultaneously in both hemispheres.

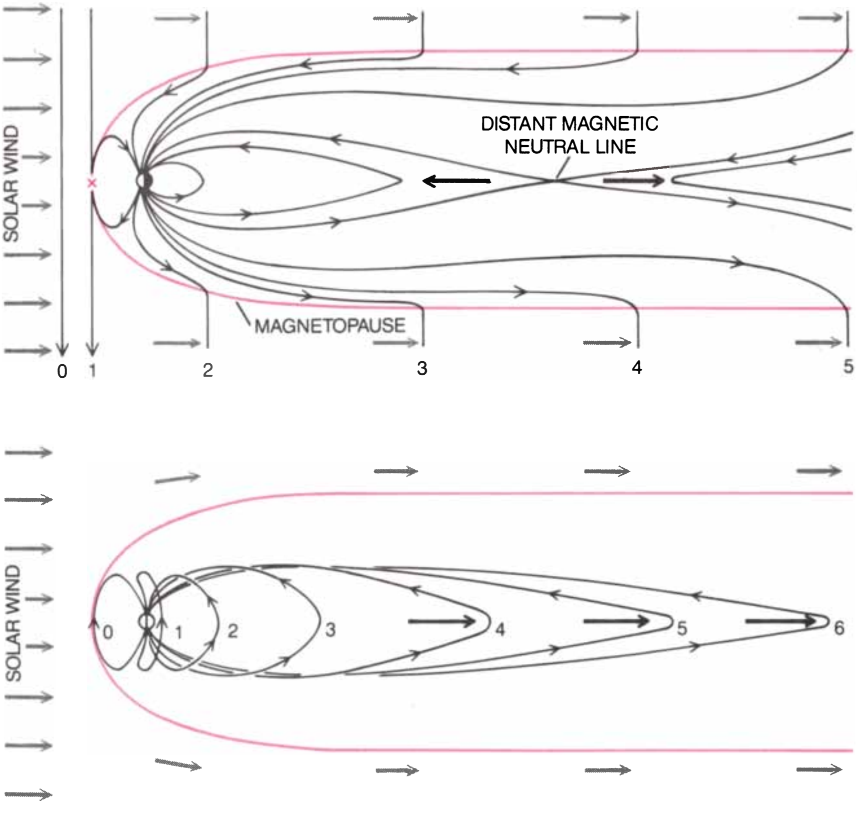
The magnetospheric phenomena of which the aurora is a terrestrial part are now at least partially understood. Indeed, by the early 1970's investigators at the University of California at Los Angeles, including Ferdinand V. Coroniti, Charles F. Kennel, Robert L. McPherron and Christopher T. Russell, had developed a theoretical model of the cause of magnetospheric substorms that remains the basis of the views widely held today. The model, now called the neutral-line model of substorms, is an outgrowth of ideas presented by James W. Dungey of the University of Cambridge in the early 1960's. My own work, which supports and extends the model, has relied heavily on magnetotail-plasma measurements made from satellites by instruments developed by Samuel J. Bame, Sidney Singer and other colleagues of mine at the Los Alamos National Laboratory. A notable aspect of the model is that it predicts the formation and expulsion of plasmoids.

At the instant an auroral substorm starts at the earth, a new magnetic neutral line, called the substorm neutral line or the near-earth neutral line, forms spontaneously in the plasma sheet about 15 earth radii downstream from the earth, or about a fourth of the distance from the earth to the orbit of the moon [*see illustration on opposite page*]. The formation of the new neutral line is usually preceded by an extreme tailward stretching of magnetic field lines at distances beyond about seven earth radii downstream from the earth. The stretching, which can develop for an hour or so, results from an increased rate of field-line reconnection at the magnetopause and hence an increased transfer of energy from the solar wind to the magnetotail by means of magnetohydrodynamics.

For its part, the formation of the new neutral line disrupts the electric current that normally crosses the plasma sheet earthward of the line. As a result this part of the cross-tail current is suddenly reduced, so that the magnetic field lines in the region suddenly become less stretched and contract or "collapse" toward the earth, becoming more dipolar in shape. The sudden collapse rains electrons into the upper atmosphere, producing the auroral lights when the electrons bombard the atmosphere's atoms some 100 kilometers above the surface of the planet.

Meanwhile the events that lead to the formation of a plasmoid have begun. The stretched field lines meeting at the new neutral line start to reconnect. They form shortened closed field lines on the earthward side of the neutral line (the field-line collapse mentioned above) and closed loops on the tailward side of the neutral line. The latter span the distance from the substorm neutral line to the presubstorm distant neutral line. The jetting of plasma from the reconnection region carries the closed loops tailward and the shortened field lines earthward at speeds of several hundred kilometers per second.

When the reconnections of closed field lines of the plasma sheet are complete, a set of nested, closed loops is free of the earth. The length of this assemblage (at the instant it forms) is from 70 to 80 earth radii. Its width may be half the tail's width, or from 20 to 25 earth radii. Its height should be the plasma sheet's height, say from 10 to 12 earth radii. This remarkable magnetic structure, with the hot plasma it confines, constitutes a plasmoid. After it forms, the surrounding open field lines in the lobes reconnect to form a sheath of interplanetary field lines, representing magnetic forces that act like a slingshot and propel the plasmoid tailward at a speed of some 500 to 1,000 kilometers per second, or about five to 10 earth radii per minute. Between the neutral line and the departing plasmoid only a very thin plasma sheet threaded by interplanetary field lines remains.



TWO SITES of magnetohydrodynamic generation arise from the interaction of the solar wind and the earth's magnetic field. Each drawing shows successive positions of a magnetic field line. In the top drawing (a vertical slice through the middle of the magnetosphere) the field line, which points southward (*0*), is initially part of the magnetic field carried outward from the sun by the solar wind. At the sunward limit of the magnetopause (*color*) the field line meets one of the earth's field lines and reconnects with it (*1*), becoming tied to the earth. The outer portion of the line is carried downstream (*2–5*) by the solar wind; the inner portion provides a path to the ionosphere through which polarization charges generated by magnetohydrodynamics can flow. In the bottom drawing (a side view of the magnetosphere) a field line at the side of the magnetosphere (*0*) is carried downstream (*1–6*) by solar-wind plasma that has penetrated the front surface of the magnetosphere. The plasma, crossing field lines such as the one in the drawing, again creates the elements of a magnetohydrodynamic generator in which polarization charges have a path to the ionosphere.

So far only about 10 minutes have passed since the onset of the substorm. An interval of from 30 minutes to one or two hours then ensues, during which the neutral line remains in its near-earth location and continues to be a site for the reconnection of field lines, forming a family of closed lines on the tailward side of the earth and continuing to cause bright auroras. The interval ends with the sudden, rapid, tailward retreat of the neutral line. Plasma jetting earthward from the retreating neutral line refills the plasma sheet, returning it to its presubstorm configuration.

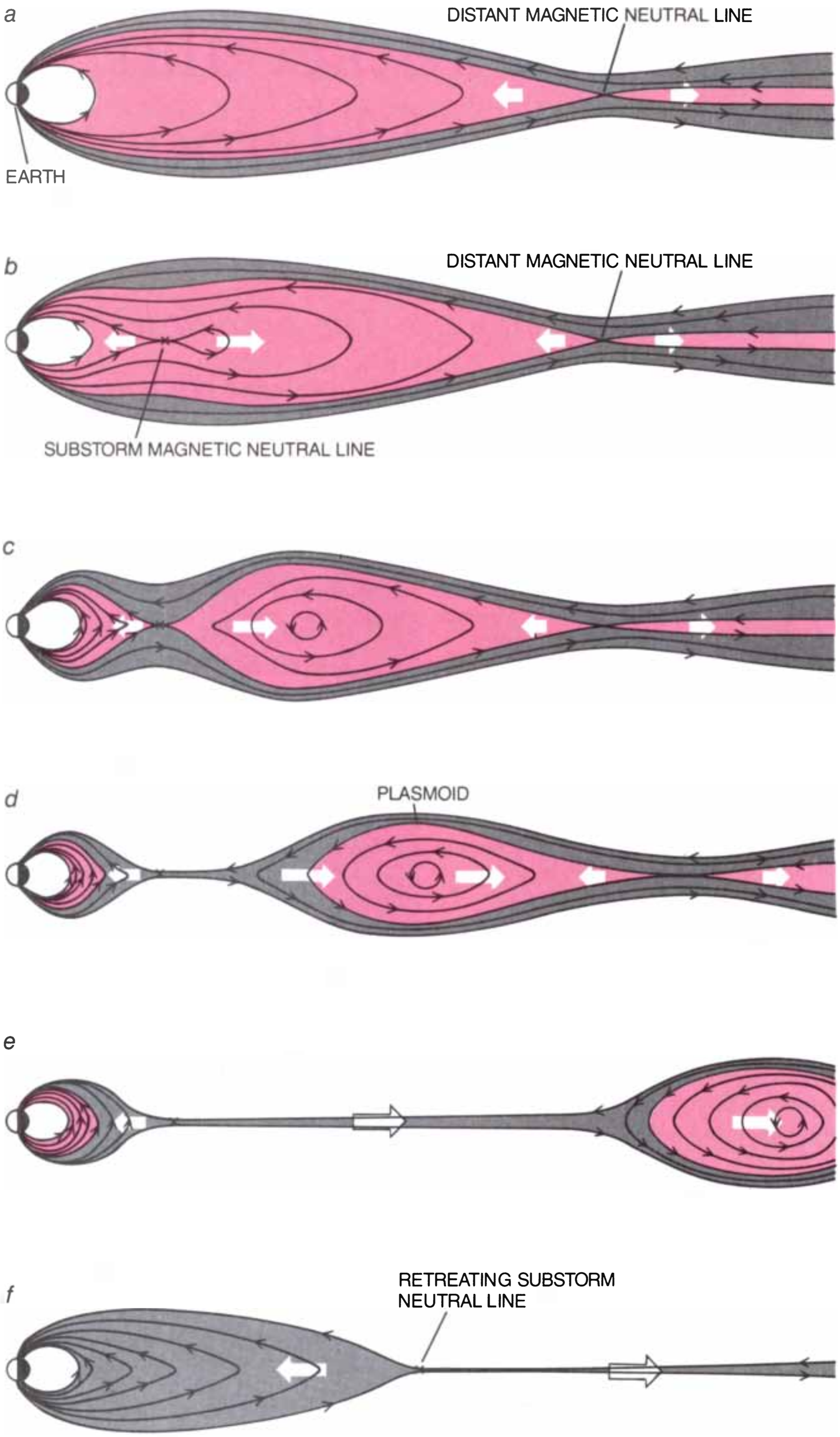
The neutral-line model of the magnetospheric events associated with a substorm derived early support from satellite observations of magnetic fields, plasmas and individual energetic particles in the part of the magnetotail extending from seven to 35 earth radii downstream from the earth. In fact, the model was developed partly as an interpretation of those observations. Then, in 1983, the satellite *International Sun-Earth Explorer 3,* or *ISEE-3,* launched by the National Aeronautics and Space Administration, made traversals of the magnetotail at distances as great as 230 earth radii. The data it returned to the earth confirmed the neutral-line model in a dramatic fashion: they revealed the passage of the plasmoids the model predicted.

The study of the earth's magnetotail has implications beyond the effort to account for electromagnetic phenomena nearby in space. Unique among plasma realms in its accessibility by satellites, the earth's magnetotail is in effect a laboratory where physical processes in astronomical plasmas can be studied in unmatched detail. Magnetic reconnection is one such process; it is probably important in many astronomical phenomena. An example is astrophysical jets: vast, narrowly collimated streams of hot plasma extending from stars, and even from galaxies. Little is yet understood of their origin or their dynamics. On the other hand, astrophysical jets are known to contain magnetic fields, and some jets contain coherent substructures whose appearance suggests that the plasma within them is magnetically confined. Conceivably these substructures are cosmological variants on plasmoids.

Magnetotails occur frequently in the solar system. They are formed wherever a body that has an intrinsic magnetic field (as in the case of Mercury, the earth, Jupiter and Saturn) or a body that has an electrically conductive atmosphere (as in the case of Venus and comets) is embedded in a flowing, magnetized plasma. Under these conditions the flow of plasma past the body is restrained by the electromagnetic forces I have described, and energy is stored in the form of stretched magnetic fields in the space downstream from the body. The intermittent release of the energy is likely to be achieved by magnetic reconnection, much as it is in the earth's magnetotail. Indeed, this explanation has been offered for disconnection events: the breaking of segments from comet tails, which is often observed.

A further step in the study of the earth's magnetotail, a step in which I am taking part, is PROMIS: the Polar Region and Outer Magnetosphere International Study. It will begin in mid-March and continue through mid-June. The idea is to coordinate the activity of European and American satellites already in space so that images of auroras made simultaneously in north and south polar regions of the earth can be related to data collected at the same time in the magnetotail, distant from the earth.

Finally, it should be noted that magnetic reconnection and plasmoid formation occur in laboratory plasmas as well as those in space. As part of the worldwide effort to harness the energy of nuclear fusion, test devices are being developed in which very hot plasma is confined in magnetic fields while its constituent atomic nuclei interact. In some recent experiments magnetic field lines threading a body of plasma are pinched off from their external source by magnetic reconnection, leaving a plasma confined in closed magnetic loops. The result is a plasmoid much like the ones the earth's magnetotail produces.


a
DISTANT MAGNETIC NEUTRAL LINE
EARTH
b
DISTANT MAGNETIC NEUTRAL LINE
SUBSTORM MAGNETIC NEUTRAL LINE
c
d
PLASMOID
e
f
RETREATING SUBSTORM NEUTRAL LINE

CREATION OF A PLASMOID is one way the magnetotail releases energy; the process, predicted by theory, was confirmed by satellite data collected in 1983. The first drawing (*a*) shows the magnetotail as it may look about an hour before the start of a magnetospheric substorm. The distant magnetic neutral line is about 100 earth radii downstream from the earth. The substorm begins when the transfer of magnetic energy from the solar wind to the magnetosphere overstretches the magnetotail's field lines and forms a substorm magnetic neutral line within the plasma sheet (*color*) about 15 earth radii downstream from the earth (*b*). Field lines reconnect rapidly at the substorm neutral line (*c*) until a plasmoid—hot plasma trapped in a nest of looping field lines—is free of magnetic attachment to the earth (*d*). Then field lines in the lobes (*gray*) begin to reconnect, forming a sheath of interplanetary field lines that acts like a slingshot, propelling the plasmoid downstream (*e*). The various reconnections also form closed field lines that contract toward the earth, injecting energetic particles into the Van Allen radiation belts and depositing in the upper atmosphere the energy that causes the substorm's auroral displays. Finally, the substorm magnetic neutral line moves downstream (*f*), reinstating the magnetotail's resting configuration.

The Molecular Genetics of Hemophilia

Hemophiliacs bleed because a defective gene deprives them of a key blood-clotting protein. The protein has now been made artificially by isolating the normal gene and then inserting it into cultured cells

by Richard M. Lawn and Gordon A. Vehar

A small defect in a single human gene, and the resulting absence or deficiency of the protein it encodes, can lead to debilitating disease. Such a disease is hemophilia. Hemophiliacs lack a crucial blood protein, one that takes part in the cascade of enzymatic reactions that causes blood to clot at the site of a wound. If a severely ill hemophiliac is not treated, he may suffer internal hemorrhaging after a minor bump; he will probably die at an early age from the effects of a bleeding crisis.

Fortunately hemophiliacs can be treated with regular transfusions of a concentrate of the missing protein. Since the early 1960's, when this form of treatment first became available, the lives of hemophiliacs in developed countries have improved dramatically, and their life expectancy, once only about 20 years, is now nearly normal. The protein concentrate, however, has to be prepared from the pooled blood of a large number of donors, and so it is expensive. In the U.S. the amount required by a typical hemophiliac in a year costs between $6,000 and $10,000; in poor countries the concentrate is often not available at all. Moreover, because it is made from pooled blood, it may spread viral diseases. Most hemophiliacs are chronically infected with hepatitis viruses, and they risk contracting the acquired immune deficiency syndrome (AIDS).

Hence there has been a strong interest in finding a way to make the antihemophilic protein by means of genetic engineering. In most cases hemophilia is caused by a defect in the stretch of DNA that encodes a clotting protein called factor VIII. Research groups at two biotechnology companies, including our own group at Genentech, Inc., in South San Francisco, have recently succeeded in isolating the factor VIII gene from the cells of healthy people and recombining it with the DNA of cells cultured in the laboratory. The recombinant cells replicate, and in so doing they make many clones, or copies, of the factor VIII gene. Each recombinant cell expresses the gene's instructions; together the cells synthesize a significant quantity of factor VIII.

The bioengineered protein works. In laboratory tests it causes blood drawn from hemophiliacs to clot, and it has proved effective in hemophilic dogs. Both Genentech and the Genetics Institute in Cambridge, the other firm involved in this research, are now developing methods for synthesizing factor VIII on a commercial scale. Although the protein must still undergo further tests in animals and then in human patients, it seems likely that within a few years abundant supplies of pure, virus-free factor VIII will be on the market.

The availability of the cloned gene is already transforming the study of hemophilia, which had been hampered by the fact that factor VIII is extremely difficult to purify from blood. In addition to being scarce, factor VIII is an unusually large and unstable protein. When we began our work, its structure was not known, nor where in the body it is synthesized. Now that it can be synthesized in the laboratory much more of it is available for study. Furthermore, the fundamental structure of the protein has been read from its genetic blueprint. In a few cases we and other workers have even been able to pinpoint the genetic mutations that give rise to hemophilia and are passed on from generation to generation.

The knowledge that hemophilia is inherited goes back at least to the writers of the Talmud: they decreed that boys whose older brothers or cousins had bled to death after circumcision need not undergo the procedure. The distinctive inheritance pattern of the disease—generally only males are afflicted, but females may be carriers—was first described accurately early in the 19th century. Perhaps the most celebrated carrier was Queen Victoria. One of her sons was hemophilic and at least two of her daughters were carriers. Through the marriages of her daughters Victoria's mutant gene spread to the royal families of Germany, Russia and Spain.

It is now known that hemophilia is sex-linked because the gene for factor VIII happens to lie on the X chromosome. Female cells contain two X chromosomes; male cells contain one X and one Y chromosome. Since a male has just one factor VIII gene, inherited from his mother, he will be hemophilic if the gene is defective. A female, in contrast, has two factor VIII genes, one inherited from each parent. She can therefore carry a defective gene without suffering from hemophilia, because the normal gene on her other X chromosome protects her. Only rarely will both genes be defective, and so there are only a handful of female hemophiliacs. Carrier females will, on the average, pass their mutant gene to half of their daughters, who will be carriers, and to half of their sons, who will be hemophilic.

The process of blood clotting that goes awry in a hemophiliac is understood only in outline. It is initiated by platelets that adhere to the site of a wound. The platelets would be easily dislodged, however, were they not bound in place by strands of fibrin, an insoluble polymer. The formation of a network of fibrin from its soluble precursor, fibrinogen, is the key event in clotting; it is the end result of a complex cascade of protein interactions that is somehow set in motion by an injury to a blood vessel. At each step in the cascade a protein precursor is cleaved to form an active enzyme called a protease. The protease thereupon cleaves another protein, converting it into a protease. Most of the cleavage steps involve cofactors, which in some cases are themselves proteins that exist in both active and inactive forms. Factor VIII, in spite of its name, is such a cofactor. It helps the protease factor IX to activate factor X in the middle of the cascade.

The clotting cascade incorporates positive-feedback loops to accelerate the clotting response and negative-feedback loops to help stop clotting. For example, thrombin, the protease that converts fibrinogen into fibrin, also activates factor VIII. At the same time, though, it activates a protease called protein *C,* which deactivates factor VIII. Since the concentration of factor VIII in normal blood is extremely small (for every molecule of factor VIII there are about a million molecules of albumin, the major blood protein), it may well be a limiting factor. In other words, the ready activation and deactivation of factor VIII may in part account for the delicate balance in healthy people between clot formation and the free flow of blood.

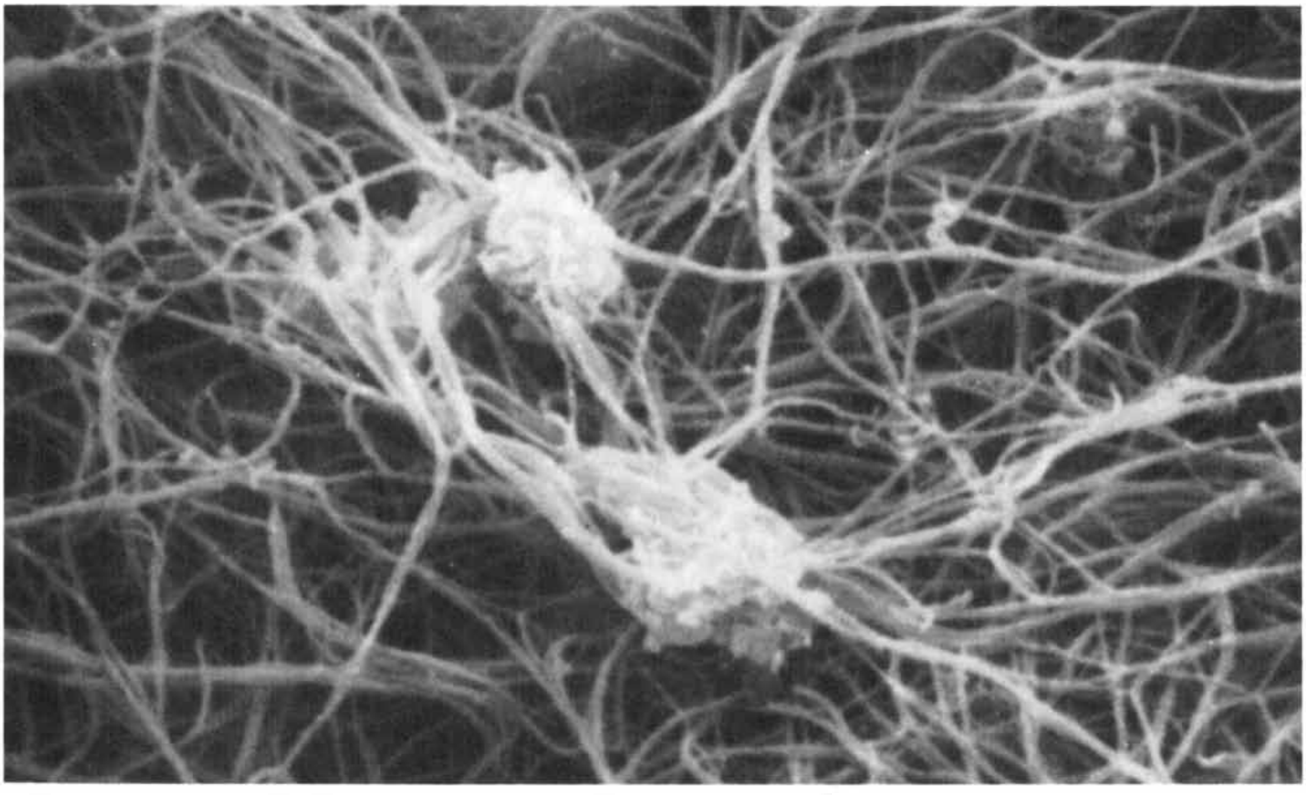
In hemophiliacs the balance is disrupted. About 85 percent of them, or roughly one male in 10,000, suffer from classic hemophilia (hemophilia *A*), in which the absence of functional factor VIII halts the clotting cascade before fibrin can form. Nearly all the rest suffer from hemophilia *B,* which is caused by a factor IX deficiency. The gene for factor IX has been cloned, and several biotechnology companies are trying to develop a bioengineered factor IX product. Bioengineered factor VIII is the greater prize, however, because hemophilia *A* afflicts more people than hemophilia *B*.

Manufacturing a protein as large and as scarce as factor VIII offered unprecedented technical challenges. The difficulties forced us to modify the standard method by which genes are cloned and manipulated to direct the synthesis of proteins.

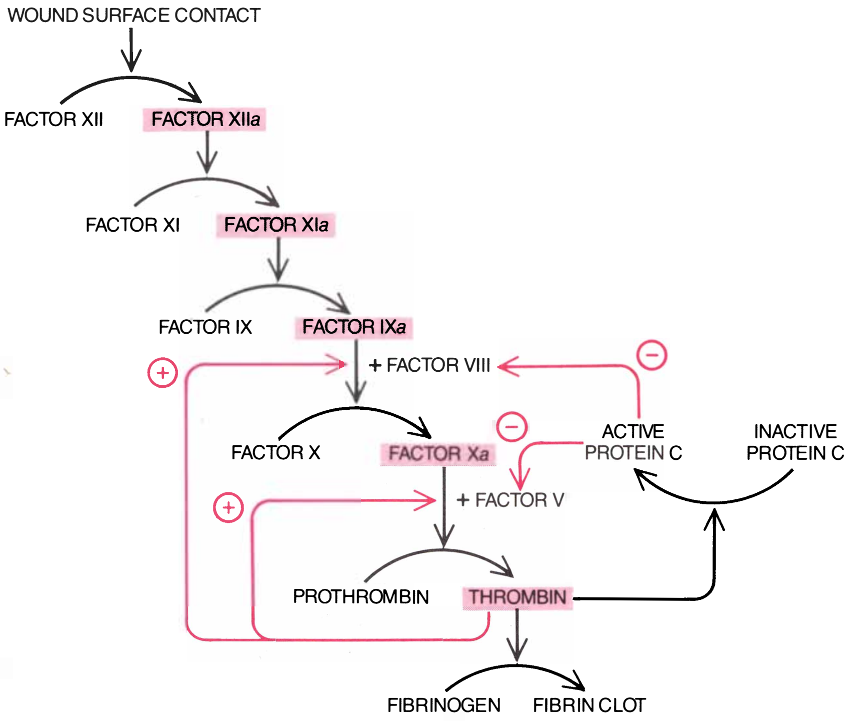
A protein such as factor VIII is a chain of amino acids; its gene is a stretch of DNA, that is, a chain of nucleotides. The sequence of amino acids is determined by the sequence of nucleotides. Each nucleotide carries one of four bases: adenine (*A*), thymine (*T*), guanine (*G*) or cytosine (*C*). A set of three bases, called a codon, specifies one amino acid. The structure of the bases is such that they form complementary pairs: adenine forms hydrogen bonds with thymine, whereas guanine binds to cytosine. Base pairing holds the two strands of the DNA double helix together. It also governs the transcription of a gene into messenger RNA (mRNA) and the subsequent translation of mRNA into protein.

In manufacturing a protein in the laboratory the key problem is to find the right gene among the thousands in a cell. Base pairing provides the solution. A small piece of DNA or RNA whose base sequence is complementary to part of the desired gene serves as a probe for the gene. A DNA probe can be made, for example, by reverse-translating part of the desired protein's amino acid sequence according to the genetic code. The synthetic probe is labeled with a radioactive nucleotide. When the probe is mixed with the DNA in a gene "library," it "hybridizes" only with the desired gene, which is thereby labeled too.

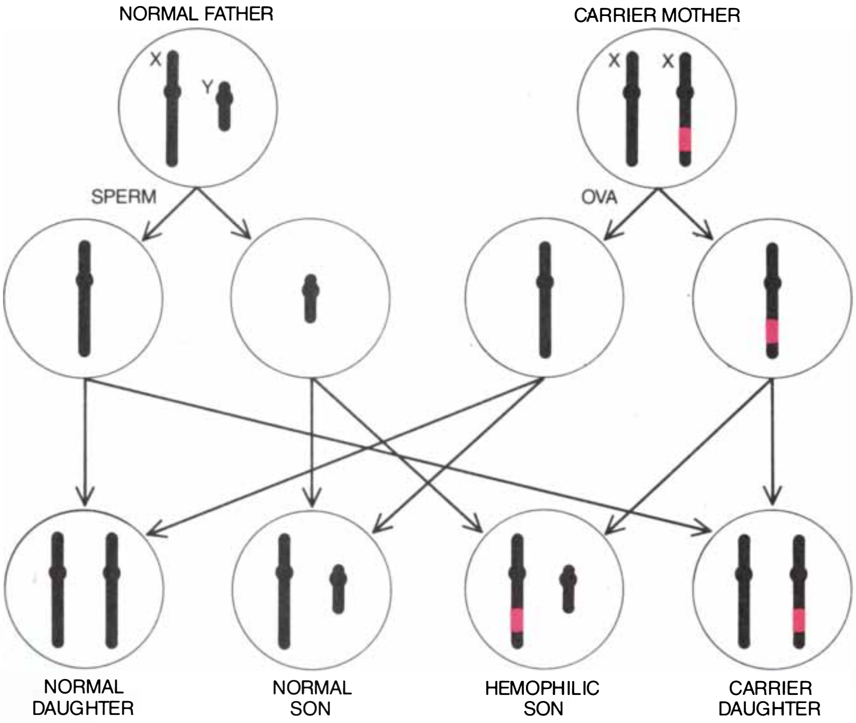
The smaller the gene library is, the easier it is to select a specific gene. The commonest cloning method, called cDNA cloning, reduces the size of the library by taking advantage of the fact that not all genes are active in every cell. In a given cell only some genes are transcribed into mRNA and then translated into protein. If it is known



FIBRIN STRANDS stabilize a blood clot at the site of a wound by trapping the platelets that form the bulk of the clot. The electron micrograph, which was made by Jon C. Lewis of Wake Forest University, shows a clot formed in a suspension of platelets and fibrin. A clot in the bloodstream is the result of a complex cascade of enzymatic reactions culminating in the conversion of fibrinogen, a soluble protein, into insoluble fibrin strands. In hemophiliacs a crucial protein in the blood-clotting cascade is either missing or defective.



CLOTTING CASCADE begins when cell damage at a wound somehow activates the enzyme factor XII; it ends with the conversion of fibrinogen into fibrin by thrombin. At each step an inactive protein is converted into a protease, or protein-cutting enzyme (*color*), which activates the next protein. Some steps require cofactors such as factors VIII and V. The cascade includes positive- and negative-feedback loops (*colored arrows*). Thrombin activates factors VIII and V; it also deactivates them (by activating protein *C*), which helps to halt clotting. Some 85 percent of hemophiliacs lack factor VIII. The rest lack factor IX.



SEX-LINKED INHERITANCE of hemophilia results from the location of the factor VIII gene on the X chromosome. A male carrying a mutant factor VIII gene lacks normal factor VIII and is hemophilic. A female carrier is protected by the normal gene on her second X chromosome, but half of her daughters will be carriers and half of her sons will be hemophilic. In the case of a hemophilic father (not shown), his sons will not be hemophilic, because they receive his Y (not his X) chromosome, but his daughters will be carriers.

which cells make the desired protein, one has only to screen the mRNA molecules from those cells. Among them there must be some transcripts of the desired gene.

To find the gene one first copies all the mRNA back into DNA with the help of an enzyme called reverse transcriptase. Individual pieces of copy DNA, or cDNA, are then enzymatically linked to the genetic material of a vector, which is often the bacterial virus phage lambda. The phages are introduced into bacteria in such a way that each phage multiplies in a separate region of a petri dish, producing a distinct plaque of phages and dead bacteria. Together the plaques constitute a cDNA library. At least one of them contains the desired cDNA fragment; that plaque is identified by hybridization with a probe.

The cDNA-cloning strategy works only if one knows what cells in the body produce the desired protein. Moreover, it is most likely to be successful if the protein is made in abundance; in that case the cells will contain many copies of the mRNA, and many plaques in the cDNA library will contain copies of the gene. Neither of these conditions applied to factor VIII. Factor VIII is scarce, and when we began our work, no one knew what organs produce it. Had we tried to construct a cDNA library, we might well have chosen the wrong cell type and ended up with a library that did not include the factor VIII gene.

We therefore decided to look for the factor VIII gene where we could be sure of finding it: in a recombinant library derived from the genome, or complete set of genes, of a cell. A genomic library is constructed by extracting the chromosomes from cells, cleaving the DNA into fragments with enzymes and joining the fragments to phage-lambda DNA. Because a genomic library contains hundreds of times more DNA than a cDNA library, it is more difficult to screen with a probe.

Before we could make a probe we first needed to know part of factor VIII's amino acid sequence. Determining even a small part of the sequence was no mean feat. The protein had not even been purified until 1980, when one of us (Vehar), working in the laboratory of Earl W. Davie of the University of Washington School of Medicine, laboriously extracted several milligrams of pure factor VIII from 25,000 liters of cows' blood. Subsequently Edward Tuddenham and his colleagues at the Royal Free Hospital in London obtained enough human factor VIII to enable workers at Genentech to sequence a short stretch of the protein. A group at the Genetics Institute achieved similar results with porcine factor VIII purified by David N. Fass at the Mayo Clinic.

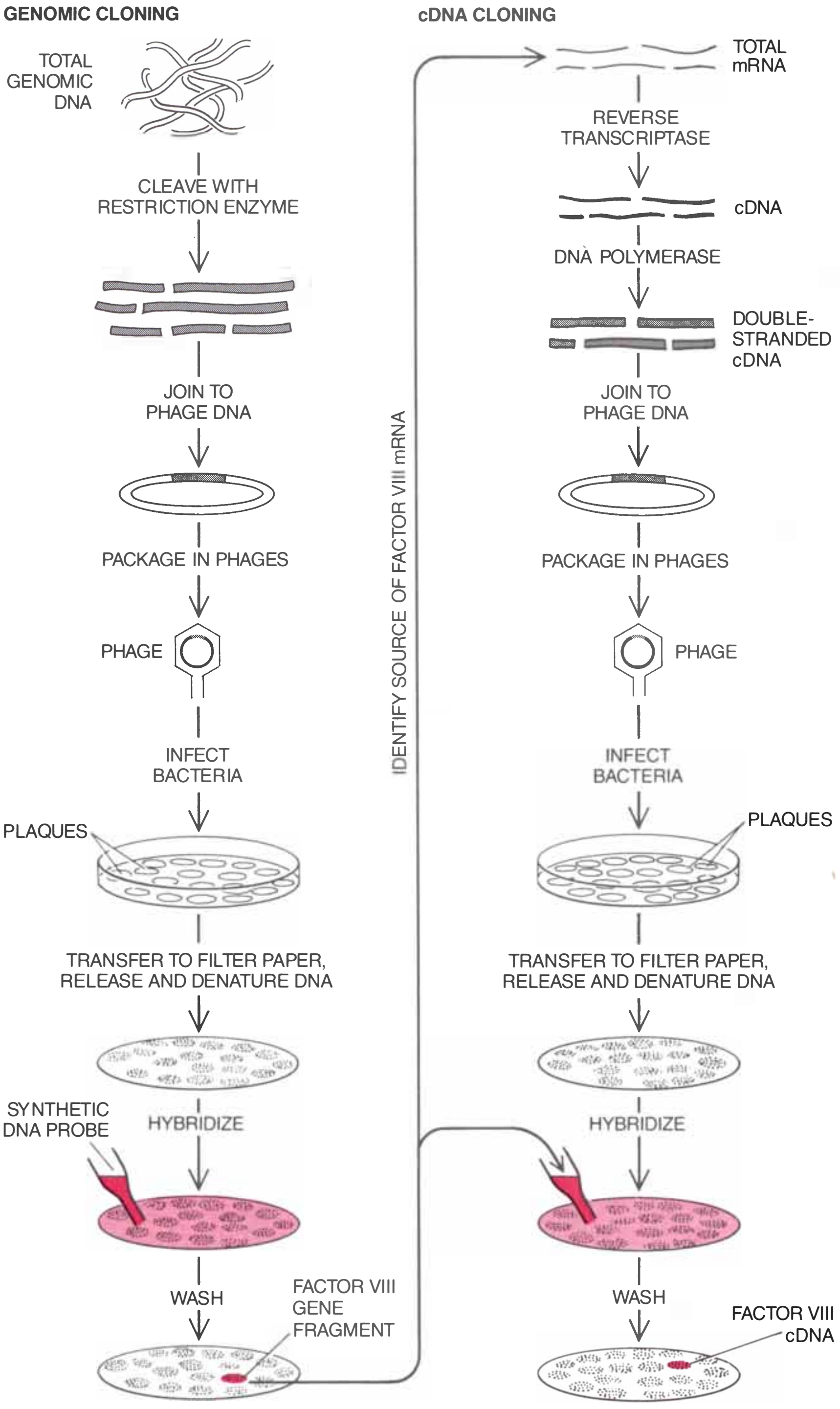
The next step was to reverse translate the protein sequence into DNA. In doing this one encounters a problem: the genetic code is redundant. An amino acid can be encoded by as many as six different codons. (There are 64 possible three-base codons but only 20 different amino acids.) One solution is to synthesize a pool of short (about 17 bases long) DNA probes that covers all the possibilities. The shorter the probe, however, the more likely it is to match randomly with a stretch of DNA other than the desired gene. When many short probes are used to screen a large genomic library, the false-positive problem becomes acute.

To avoid the problem we relied on a single, relatively long (36 bases) probe derived from a 12-amino-acid stretch of factor VIII. We had to choose among 147,456 ways of encoding this particular sequence. Fortunately we were able to make highly educated guesses, because we knew that some codons are more prevalent than others in mammalian genes. As it turned out, we got 30 of the 36 bases right, which was a close enough match. When we screened the genomic library, the synthetic probe hybridized with overlapping segments of factor VIII DNA, thereby identifying plaques containing parts of the gene.

The entire gene is too large to fit into a single phage. To find the rest of it, William I. Wood, Jane Gitschier and other workers in our laboratory rescreened the library, this time using fragments of the identified gene segments as probes. By repeating this procedure, known as chromosome walking, they eventually obtained a series of overlapping segments constituting the complete gene sequence.

The gene is 186,000 bases long. The information for factor VIII, however, is spread among 26 exons, or coding sequences, that together account for less than a twentieth of the total length of the gene. The reason is that the exons are separated by 25 introns, or noncoding intervening sequences. After the entire gene is transcribed into RNA in a living cell, the introns are cut out of the transcript. The exons are then spliced to form the mRNA that directs protein synthesis. To make factor VIII in cultured cells, we too needed a gene without introns. In other words, we had to find factor VIII mRNA and convert it into cDNA.

With pieces of the real gene available as probes it became a straightforward task to find out what cells make factor VIII and its mRNA. If



GENE CLONING involves finding a specific gene among thousands in a human cell. The standard method, if one knows which cells make the desired protein, is to screen a copy DNA (cDNA) library derived by reverse transcription from the messenger RNA (mRNA) of those cells (*right*). In looking for the factor VIII gene, however, the authors did not know where the protein is produced. Hence they screened the entire human genome (*left*). Chromosomal DNA fragments were joined to the DNA of the bacterial virus phage lambda. Each phage contained one human DNA fragment; each phage multiplied and formed a plaque in a distinct region of a bacterial culture. To identify the plaque containing the factor VIII gene, the phages were blotted onto filter paper and broken open to release their DNA. The DNA was exposed to a radioactive probe: a small piece of synthetic DNA encoding part of factor VIII. The probe hybridized with part of the factor VIII gene, thereby labeling it. To produce factor VIII in cultured cells, it was still necessary to make factor VIII cDNA, which lacks the introns (noncoding sequences) that complicate the full gene. Now fragments of the cloned gene could serve as reliable probes, first for identifying cells that make factor VIII mRNA and then for finding factor VIII cDNA in the cDNA library.

the mRNA from a particular cell type failed to hybridize with the probe, one could now be sure it was not because the probe was faulty; with the synthetic probe there would always have been uncertainty. A group led by Daniel Capon at Genentech found tiny amounts of hybridizing mRNA in the mRNA of a cultured cell line, while John J. Toole and his colleagues at the Genetics Institute identified liver cells as a source. Since then factor VIII mRNA has also been found in other tissues, including kidney, spleen and lymph cells, but the liver seems to be the primary source. Most of the factor VIII in healthy people is probably synthesized in liver cells and secreted into the bloodstream.

Once a source of factor VIII mRNA had been found, workers at both companies constructed cDNA libraries. Again pieces of the cloned gene served as probes, this time to pick the rare factor VIII cDNA out of a library of thousands of clones. Actually the factor VIII cDNA had to be stitched together from several overlapping segments; the mRNA is about 9,000 bases long, and with current techniques it is not possible to copy so large a molecule in one stretch. Finally, control sequences had to be attached to the cDNA. Control sequences direct enzymes in the recombinant cell to start and stop transcribing a gene.

Recombinant bacteria, usually *Escherichia coli,* were suitable for manufacturing the first bioengineered proteins, such as insulin and interferon, because these molecules are relatively small. Bacteria are generally not equipped, however, to produce a large and complex protein such as factor VIII. Nor do they always have the right enzymes for modifying or folding a large protein after it has been synthesized.

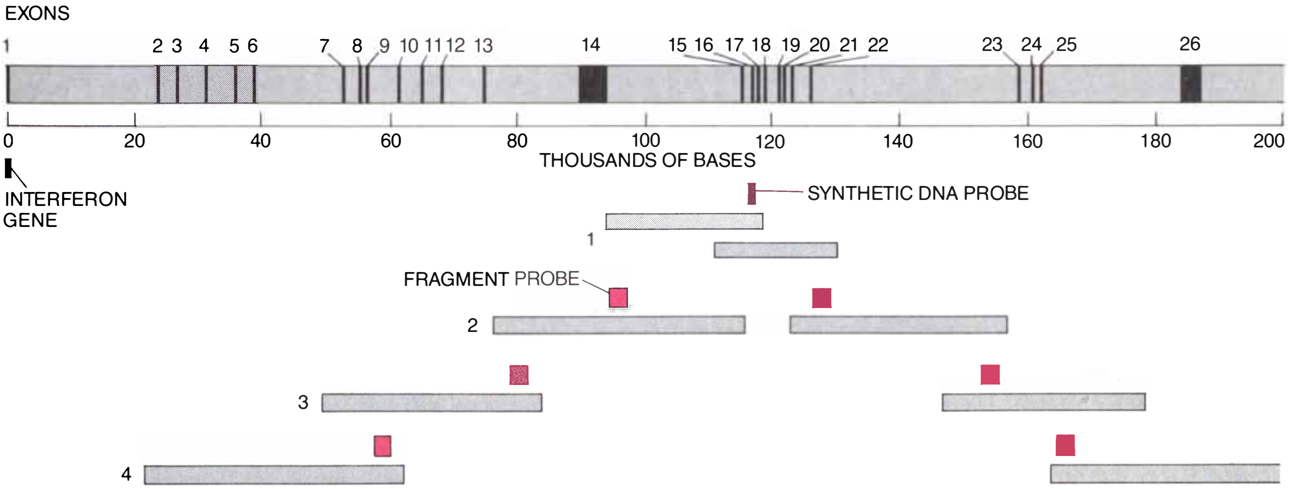
For these reasons we chose to insert the cloned gene for factor VIII into hamster cells, which are easily grown in the laboratory. The recombinant hamster cells thereupon secreted human factor VIII into the culture broth. That is, we hoped it was factor VIII; we could not be sure the protein was fully functional until it had been shown to make blood from hemophiliacs clot normally. It was possible, for instance, that our bioengineered protein was just one part of a protein complex that is missing in hemophiliacs. It was even conceivable, given how difficult it had been to purify factor VIII, that we had cloned a gene encoding some impurity in the preparation. Such unsettling possibilities had to be considered, but neither proved to be true. Bioengineered factor VIII does clot hemophilic blood. Indeed, it has been found to be equivalent in every way to the blood-derived protein.

What is the structure of factor VIII? The advent of gene cloning has produced the novel situation in which some protein sequences are determined indirectly, from the DNA sequence of the gene. Such was the case with factor VIII, which is much too large and scarce to be sequenced directly in its entirety. Before its gene was cloned workers did not even agree on its approximate size; estimates differed by a factor of nearly 100.

Now the question can be answered. The 9,000-nucleotide cDNA for factor VIII encodes a protein 2,351 amino acids long. (Nearly 2,000 bases at the ends of the gene are transcribed into mRNA but are not translated into protein.) The first 19 amino acids form a hydrophobic sequence typical of secreted proteins. This "signal peptide" is generally cut off the protein as it is secreted, and so a mature factor VIII molecule must consist of the remaining 2,332 amino acids. Its molecular weight must be about 330,000 daltons. (One dalton is 1.66×10^{-24} grams.) In comparison, a molecule of interferon is only 166 amino acids long and weighs about 19,000 daltons. The factor VIII gene is by far the largest ever cloned and expressed in foreign cells.

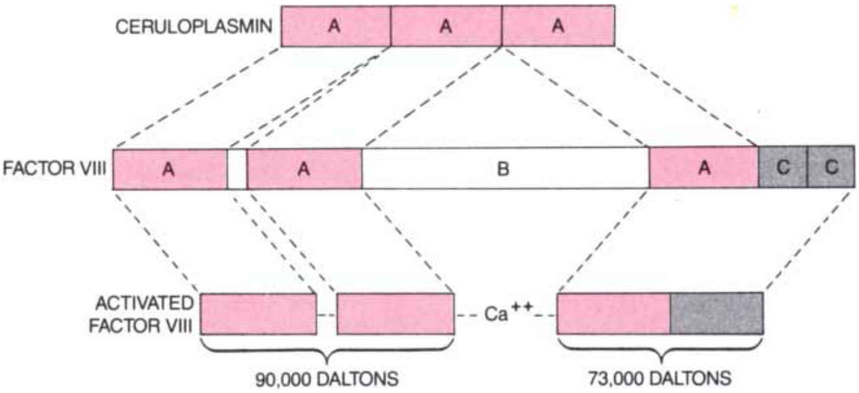
An analysis of the amino acid sequence of factor VIII reveals that the protein is constructed of repeated similar segments. Three of these are designated *A*. Of the roughly 350 amino acids that make up the sequence of each *A* segment, approximately a third are common to all three segments. A comparable level of homology also exists between the two 150-amino-acid segments designated *C,* which are not homologous to the *A* segments. Since there are 20 different amino acids, a one-third homology among sequences is almost certainly not a random occurrence; the homologous segments must be related.

A surprising clue to the evolutionary history of factor VIII came from a computer-aided comparison of its sequence with that of other proteins. The three *A* segments turn out to be nearly



TREMENDOUS SIZE of the factor VIII gene, the largest gene cloned to date, forced workers to apply a cloning technique called chromosome walking. The factor VIII gene is 186,000 bases long. In contrast the interferon gene, which was cloned in 1980, incorporates only about 600 bases. Because the factor VIII gene is too large to fit into a single phage, segments of it were found in different plaques in the genomic library. When the library was screened with a synthetic DNA probe, the probe hybridized with overlapping segments (*1*). Pieces of the segments then served as probes to rescreen the library and identify further segments (*2*). By repeating this procedure nearly all of the gene was identified (*3, 4*). (Its beginning was found once factor VIII cDNA was available as a probe.) Less than one-twentieth of the gene consists of exons, or coding sequences (*black bands*); the 26 exons are separated by 25 introns.

as similar to the three domains of ceruloplasmin, a protein that carries copper in the bloodstream, as they are to one another. Previously there had been no reason to suspect a connection between factor VIII and ceruloplasmin, but the homology of their domains suggests their genes evolved from a common ancestor. The ancestral protein may have consisted of three identical domains, precursors of the modern *A* domains. In ceruloplasmin the three *A* domains are contiguous and make up the entire molecule, whereas in factor VIII the second and third *A* domains are separated by nearly 1,000 amino acids. In the factor VIII gene this intermediate region is encoded by a single huge exon, which may have been inserted into the ancestral gene. The regions encoding the *C* segments of factor VIII were also added to the end of the gene.



AMINO ACID SEQUENCE of factor VIII resembles that of the blood protein ceruloplasmin. The three *A* domains of factor VIII have about a third of their amino acids in common and are also homologous to the three domains of ceruloplasmin. The two proteins must have evolved from a common ancestor. Factor VIII probably diverged from ceruloplasmin when a large exon (number 14) encoding the intermediate *B* segment was inserted into the ancestral gene; exons encoding the *C* segments were added to the end. When factor VIII is activated, the *B* segment is excised by proteases, and the two resulting subunits are held together by a calcium ion. Another cleavage takes place between the first two *A* domains.

Factor VIII purified from donated blood is almost never identical with the full, 330,000-dalton molecule encoded by the gene. It is now thought that factor VIII undergoes a series of cleavages, including the removal of the region between the second and third *A* domains (the *B* region), when it is activated in the bloodstream. Evidence for this view was found by sequencing small parts of the active protein directly and comparing the results with the sequence of the cloned gene. Active factor VIII seems to consist of a 90,000-dalton subunit joined to a 73,000-dalton subunit. The first subunit consists of the first two *A* domains; the peptide bond between the two is cut but they remain linked. The second subunit consists of the third *A* domain and the two *C* domains. Just as the molecule is activated by cleavages, so too is it readily deactivated by further cleavages within the subunits. It is probably through these reactions that the clotting cascade is brought to a timely halt.

The details of how factor VIII functions in the cascade are far from clear. It is bound to a carrier protein called von Willebrand factor, which keeps factor VIII circulating in the blood and may help to position it on the surface of a platelet at the site of a wound. Once on the platelet, factor VIII probably separates from von Willebrand factor and forms a complex with factor IX and factor X. The binding sites on these proteins have not been found. All that is really known is that without factor VIII the activation of factor X by factor IX does not take place.

A standard method of learning more about how a protein works is to study the abnormal forms that result from genetic mutations. We can now begin to apply this approach to factor VIII and at the same time identify the mutations that cause hemophilia. The ultimate result will be an improved understanding of the disease and improved treatment for its victims.

As expected, we and other workers have found there is no single mutation underlying hemophilia. Fifty years ago the British geneticist J. B. S. Haldane pointed out that serious diseases linked to the X chromosome must constantly arise anew through random, spontaneous mutations; otherwise the diseases would eventually die out. Indeed, roughly a third of the cases of hemophilia observed today occur in families with no history of the disease. A hemophilic gene is of course transmitted to offspring, but before the advent of effective treatment a particular mutation soon became extinct, simply because there were fewer surviving children in hemophilic families. In contrast, a recessive mutation on an autosomal chromosome, of which a cell has two copies, can spread through the population, because it affects only those rare individuals who inherit two defective genes.

In principle it is possible to identify the mutation that gives rise to hemophilia in an individual by isolating and sequencing his factor VIII gene. Since it would take several months to sequence each 186,000-base gene, however, the method is not practical. Fortunately there is a quicker procedure, albeit one that is applicable to only a small set of cases. The procedure is based on a hybridization technique called Southern blotting.

The first step is to extract the DNA from the blood cells of a hemophiliac. The DNA is cleaved into a million or so fragments with a restriction endonuclease, an enzyme that cuts DNA wherever it recognizes a specific four-to-six-base sequence. The fragments are then separated according to size by electrophoresis: the smaller the fragment is, the farther it migrates through an agarose gel when an electric current is applied to the gel. Next the DNA is unraveled into single strands and blotted onto special filter paper. In the process the fragments retain the relative positions they occupied in the gel. Finally the filter is bathed in a solution of radioactive factor VIII cDNA. The cDNA probe hybridizes with fragments of the factor VIII gene. The sizes of the fragments can be deduced from their positions on the paper, which form a distinctive pattern.

To find factor VIII mutations, we and our colleagues compared the hybridization patterns of normal and hemophilic DNA's. Two types of gene alteration can be detected with this method. The easiest to recognize is a gross deletion of part of the factor VIII gene; some of the hybridizing fragments are missing or are altered in size. Occasionally, though, one can also detect a change of a single DNA base, provided the change happens to occur within the recognition sequence of a restriction enzyme. Such a mutation prevents the enzyme from cleaving the gene. Two hybridizing fragments in the normal pattern are therefore replaced by a single larger fragment in the hemophilic pattern. (Conversely, a mutation can also change a blot pattern by creating a new restriction site.)

One example of a single-base mutation will suffice to illustrate the procedure. DNA from a severe hemophiliac, from his parents and from his three siblings was cut with the restriction enzyme *Taq*I, which recognizes the base sequence *TCGA*. The Southern blots of the five unaffected relatives all showed two hybridizing fragments, one consisting of 1,400 nucleotides and the other of 2,800 nucleotides. On the hemophiliac's Southern blot the two fragments were replaced by a single fragment 4,200 nucleotides long. Because we knew the sequence of the normal factor VIII gene, we could determine the position of the altered *Taq*I site. By cloning and sequencing that part of the hemophiliac's gene we found the DNA sequence of the *Taq*I site had been changed from *TCGA* to *TTGA*. The mutation prevents *Taq*I from cleaving the gene at that point.

More important—and coincidentally—the mutation changes a codon for the amino acid arginine (*CGA*) to a "stop" codon (*TGA*), which brings the synthesis of factor VIII to a premature halt. The truncated protein is probably either inactive or too unstable to survive in the bloodstream. Since the hemophiliac's parents lack the mutation, he did not inherit it. It must be a new mutation that occurred in the egg from which his cells developed.

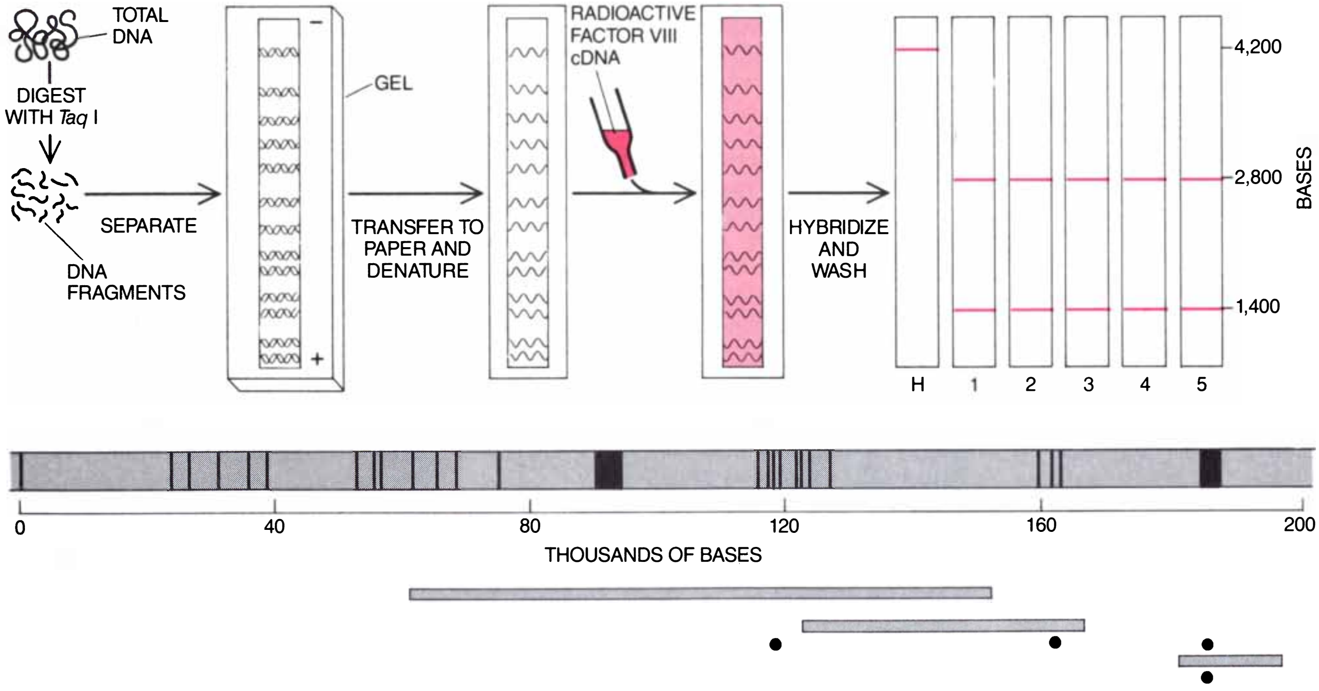
So far a total of 200 hemophilic factor VIII genes have been examined in our laboratory and by workers at the Genetics Institute and at Johns Hopkins University. Seven different mutations have been pinpointed, and none has been observed in more than one family. Four of the mutations are single-base changes, of which three lead to a truncated factor VIII and severe hemophilia; the fourth causes the substitution of an incorrect amino acid and results in a relatively mild form of the disease. The other three mutations are deletions of several thousand nucleotides from the gene. All three deletions cause severe hemophilia.

In the future investigators may have access to more efficient techniques for locating single-base mutations. By analyzing a large number of mutations one may be able to correlate types of mutation with the level of clinical severity of the disease. It would be of particular importance, for example, to understand why some 10 percent of hemophiliacs suffer immune reactions to exogenous factor VIII; these people are the hardest to treat.

In principle it would be possible to cure hemophiliacs by introducing functional factor VIII genes into their cells. Yet gene therapy for any disease is probably years away. One of the chief obstacles is the problem of controlling the productivity of the inserted genes: too much factor VIII, for example, may be as dangerous as too little.

The cloned factor VIII gene is already serving as the basis for more reliable methods of diagnosing female carriers and of detecting hemophilia in fetuses. Essentially these methods involve blot hybridization tests using the cloned gene as a probe to track the inheritance of a defective gene. The prenatal-diagnosis technique is being practiced at some 70 medical centers around the world. Although it is not yet applicable in all cases, it is more reliable than the old method of measuring the concentration of factor VIII in fetal blood, and it does not require an incision. In addition, whereas a fetal blood test cannot be done before the 20th week of pregnancy, DNA-based diagnosis is feasible in the eighth week. If the parents choose an abortion, it is less risky for the mother at that stage.

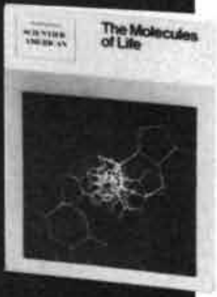
Clearly the most significant immediate medical implication of the cloning of the factor VIII gene is the prospect of a safe and abundant supply of factor VIII. The bioengineered protein is scheduled to begin several years of clinical trials within a year or two. When it becomes commercially available, hemophiliacs will be liberated from the menace of transfused infection. Many of those who live in underdeveloped countries, and who today still die young, will receive effective treatment for the first time.



HEMOPHILIA-CAUSING MUTATIONS in the factor VIII gene can be detected by Southern blotting (*top*) if they happen to change the way the gene is fragmented by a restriction enzyme. DNA from blood cells is cut into millions of fragments, in this case with the enzyme *Taq*I. The fragments are separated according to size by electrophoresis, unraveled into single strands and blotted onto filter paper. The filter is bathed in a solution of radioactive factor VIII cDNA, which hybridizes only with fragments of the factor VIII gene. The size of the hybridizing fragments is revealed by exposing X-ray film to the filter. In the example shown here a point mutation in the factor VIII gene of a hemophiliac (*H*) has eliminated a *Taq*I cleavage site. The 2,800- and 1,400-base fragments on the blot patterns of his relatives (*1–5*) are replaced by a single, uncut 4,200-base fragment. So far seven different mutations have been located on hemophilic factor VIII genes (*bottom*). Four are point mutations, or changes of a single base (*dots*); three are extensive deletions (*bars*).

THE BEST OF SCIENTIFIC AMERICAN

Selected, Arranged, and Introduced
By Leading Scientists In Attractive,
Special-Topic Book Editions



THE MOLECULES OF LIFE

Readings from **Scientific American**

Eleven world-famous scientists examine the basics of the most profound intellectual revolution of our era—the exploration of the fundamental mechanisms of life. Advances in this amazing new field have led to such formidable prospects as the correcting of genetic defects by reprogramming human genes and the production of artificial antibodies, rare hormones, and potent—yet safe—vaccines. With its outstanding illustrations and clear, succinct writing, THE MOLECULES OF LIFE provides readers with access to the most authoritative accounts of what this new science has achieved and the promise it holds for tomorrow.

1986	139 pages	101 illustrations
ISBN: 0-7167-1783-2		$12.95

CANCER BIOLOGY

Readings from **Scientific American**

With Introductions by Errol C. Friedberg, STANFORD UNIVERSITY SCHOOL OF MEDICINE

Offering a solid grounding in the causes and development of cancer, CANCER BIOLOGY consists of 13 *Scientific American* articles—selected and arranged into 4 thematic sections—that emphasize current issues as well as essential historical perspectives. Dr. Friedberg's introductions to each section include a definition and classification of the disease, an examination of known or suspected carcinogenic agents, and a scrutiny of events that perpetuate the survival and proliferation of cancer cells. For lay readers as well as for workers in the field, CANCER BIOLOGY will prove essential to the development of a comprehensive awareness of the vocabulary and concepts that underlie achievements in this increasingly significant discipline.

1985	156 pages	133 illustrations
ISBN: 0-7167-1751-4		$12.95

THE MIND'S EYE

Readings from **Scientific American**

With Introductions by Jeremy M. Wolfe, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

An entertaining and highly rewarding journey through the visual system, THE MIND'S EYE leads the reader from a look at different kinds of eyes, through the early stages of visual information processing, to the mind's creation of visual experience, including illusions and hallucinations. THE MIND'S EYE, examining the hidden, unconscious, and automatic processes that generate visual perception, is a collection of 12 *Scientific American* articles selected and organized into *thematic* sections, each of which is introduced by Professor Wolfe to elucidate one of the three basic sequential phases of vision: *Reception, Extraction,* and *Inference.*

1986	127 pages	139 illustrations
ISBN: 0-7167-1754-9		$11.95

LANGUAGE, WRITING, AND THE COMPUTER

Readings from **Scientific American**

With Introductions by William S-Y. Wang, UNIVERSITY OF CALIFORNIA AT BERKELEY

From the inception of speech through the advent of writing to the fascinating man/machine language interface of the computer age, LANGUAGE, WRITING, AND THE COMPUTER strikingly illustrates some of the most influential events in the course of human biological and cultural evolution. Together with Dr. Wang's Introduction, the 11 articles that make up this volume reveal instances of how meaning was first paired with sound to make speech, how script ranges from symbols representing meaning to those representing both meaning and sound, and how research in artificial intelligence is yielding valuable insights into the ability exercised by "human" intelligence in dealing with ambiguity.

1986	124 pages	108 illustrations
ISBN: 0-7167-1772-7		$12.95

Available at fine bookstores, or order directly from the publisher. (When ordering by mail, please add $1.50 for postage and handling. CA, UT, and NY residents please add sales tax.)

W. H. FREEMAN AND COMPANY
4419 West 1980 South
Salt Lake City, Utah 84104

THE NUTS AND BOLTS OF TODAY'S GM.

Lasers. Electronics. Polymers.
Computers. Robotics.
The new tools of a new technology.
The changing parts and components
of a General Motors which is building
a whole new future of transportation.
We're scrapping the old. Inventing the new.
Assembling a GM structured around
creative and searching minds
in research, design, manufacturing,
and engineering.
At GM today, we know that you can't
change the future
without changing
the way
you do things.

1.

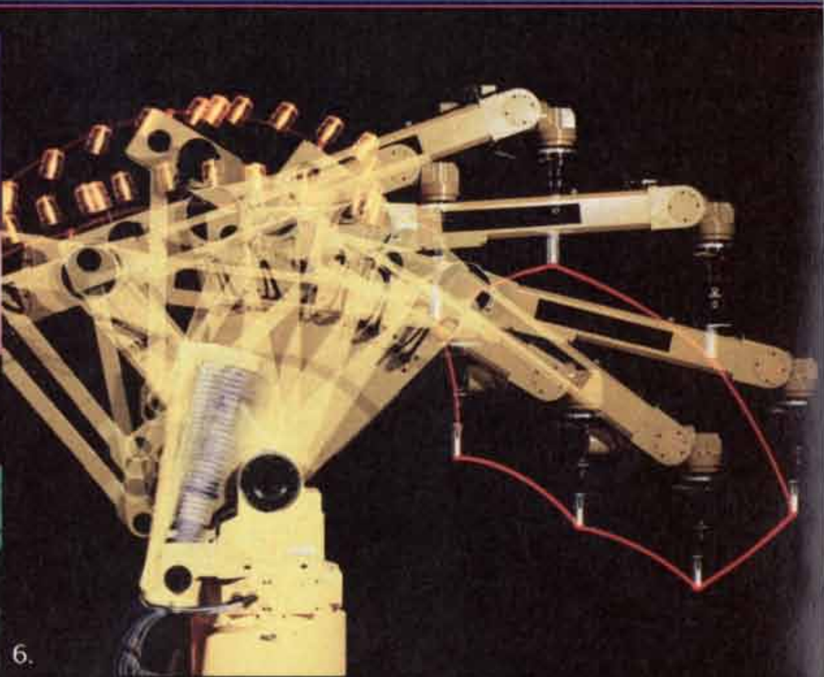
2.

THE GM ODYSSEY: SCI

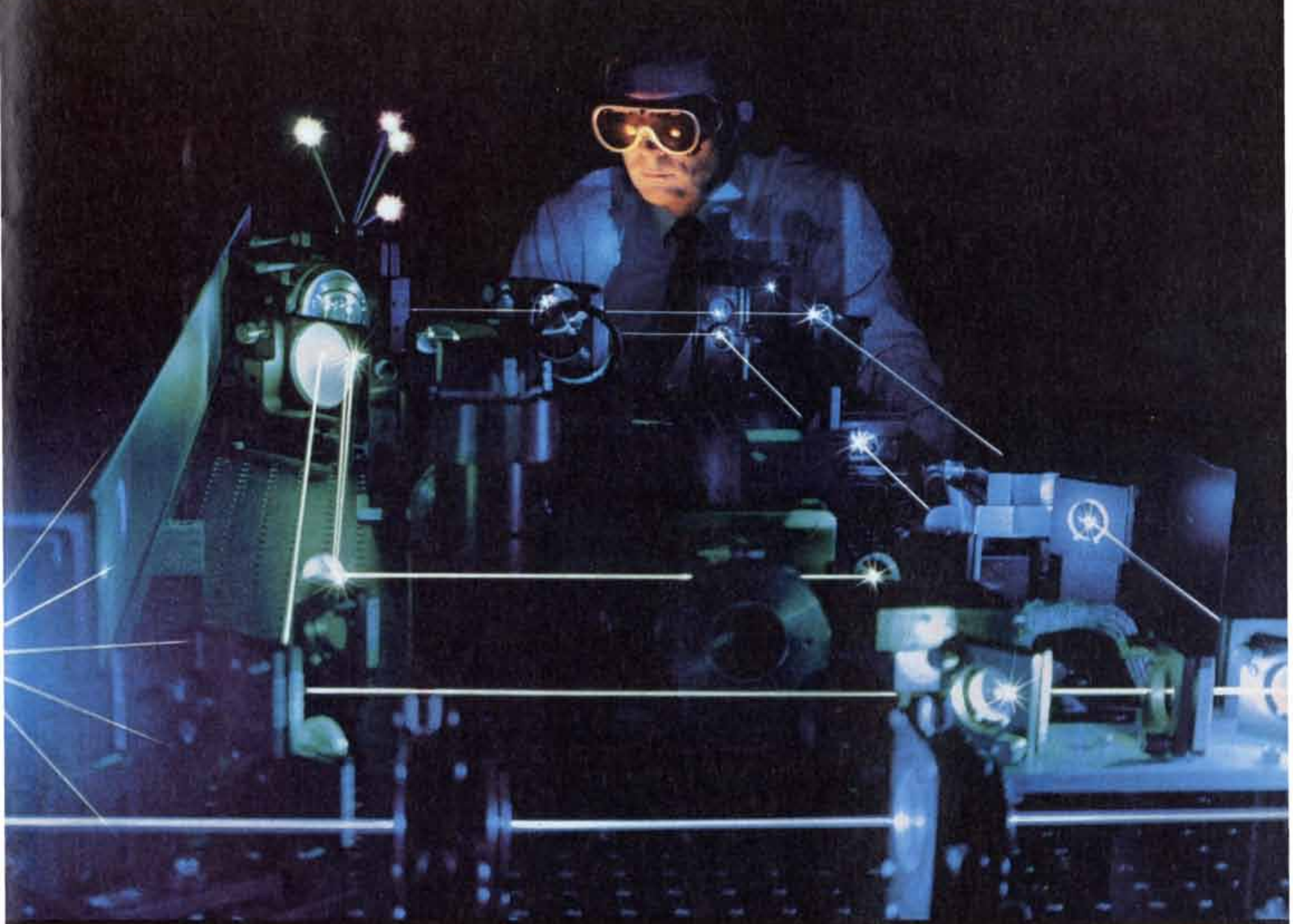
3.

4.

5.



6.



ENCE NOT FICTION

GM

1. ELECTRONIC NAVIGATION. Electronic navigation system that allows you to find where you are and how to get where you're going.
2. LASER DOPPLER VELOCIMETRY. Advanced use of lasers to develop more powerful, fuel efficient, and cleaner burning engines.
3. PLASMETAL. New plastic plating process that provides a durable finish on chrome-plated parts with no adverse environmental impact.
4. MICROELECTRONICS RESEARCH. Discovering new ways to make high performance integrated circuits for use in on-board computers, radios, and engine controls.
5. EXERCISER. Testing system used in the vehicle development phase to assure performance of on-board electronics.
6. MINTIME CONTROL. Quantum leap in controlling robotic accuracy in placement of parts and welds leading to higher quality assembly.
7. AUTOCOLOR. Computer graphics system that displays 3-D color shaded vehicle designs for aesthetic evaluation and improved styling.

7.

SCIENCE AND THE CITIZEN

Marketing SDI

The Reagan Administration is apparently trying to reposition the Strategic Defense Initiative (SDI). As the original goal of total population protection recedes into the indefinite future a new justification has begun to take shape. "The Strategic Defense Initiative is a prudent and necessary response to the ongoing extensive Soviet anti-ballistic-missile effort," Secretary of Defense Caspar W. Weinberger and Secretary of State George P. Shultz write in the preface to a Government publication titled *Soviet Strategic Defense Programs.* Yet an examination of the publication and of other information released by the U.S. Government raises the question of whether the Soviet program calls for anything like the heaven-storming, gigabuck effort being advocated by the Administration.

Department of Defense reports indicate that although advances made by the U.S.S.R. in such conventional technologies as radar and interceptor missiles would seem to rival advances made by the U.S., the U.S.S.R. lags in the development of many critical and even essential new technologies. For example, it is reported that the U.S.S.R. is particularly weak in computer and sensor technologies, two areas that are crucial to the raising of an effective shield against missiles.

According to *Soviet Strategic Defense Programs,* the Soviet Union has the only operational anti-ballistic-missile (ABM) system in the world. The system, known as GALOSH, surrounds Moscow. It consists of a network of radars known as hen houses sited along the periphery of the U.S.S.R. and two phased-array radars called the dog house and cat house near Moscow. In spite of the apparent impressiveness of the system, it does have weaknesses. Eric Stubbs of Harvard University, an analyst for the Council on Economic Priorities, writes in a report published by the council that the size of the radars makes them particularly vulnerable to attack. Although the U.S. is allowed such a system by the ABM treaty, the Nixon Administration conceded that it could be overwhelmed by a large enough attacking force.

From a technological point of view the U.S.S.R.'s ballistic-missile-defense (BMD) system is not a state-of-the-art affair. In discussing it, Sayre Stevens, the former Deputy Director of Intelligence for the Central Intelligence Agency, notes: "While the Soviet BMD program has momentum and has made significant technological progress over the past decade, it really has only reached the level of technology that was available to the U.S. ten years ago. The major difference is that the Soviet technology is much closer to application."

Soviet progress in the new "Star Wars" technologies appears to be uneven. Although the U.S.S.R. has made significant advances in laser weapons, according to Government reports, the development of particle beams and kinetic-energy weapons (chemical guns and electromagnetic guns) remains in its infancy. Indeed, an assessment prepared by the Department of Defense Program for Research, Development, and Acquisition indicates that the U.S.S.R. does not lead the U.S. in any basic strategic technologies. The assessment holds that whereas the two countries are equal in five areas, including directed-energy weapons (lasers and particle beams) and power sources, the U.S. leads the U.S.S.R. in 16 areas. They include computer hardware and software, electro-optical sensors, optics and propulsion. For example, a current Soviet computer, model ES 1060, was first produced in 1978 and is roughly equivalent to the IBM 360 computer introduced in the U.S. in the mid-1960's.

An additional handicap facing the Soviet strategic defense is that the U.S.S.R. has not yet gathered its programs together under a central management, such as the Strategic Defense Initiative Organization (SDIO). The situation in the U.S.S.R. could change. As Stubbs notes, further development of the "Star Wars" program in the U.S. could stimulate the Soviets to upgrade their technological capabilities "more aggressively."

Such a possibility, as well as a comparison of superpower capabilities, may provide a reason for the U.S. to move ahead at a more measured pace in the fields of research gathered under the SDIO's control. After all, research in lasers, particle beams, sensor technology and computer technology as well as in related fields would proceed with or without the SDI. If it does so at a normal rate instead of at the forced march promoted by SDIO supporters, such a program would have merits worth considering. According to some experts, the effort could be kept securely within the bounds of the ABM treaty and yet be robust enough to provide confidence that the Soviets would not be able to achieve a breakout.

Equally important, such a program would not threaten the U.S.S.R. with the prospect that the U.S. is moving rapidly toward a situation in which this country could, from behind an impenetrable shield, threaten nuclear blackmail or a first strike. The U.S. has assured the Soviets that no thoughts could be further from its strategic mind. Unfortunately, as some experts in Soviet affairs point out, Kremlin planners—like their counterparts in the Pentagon—would probably feel bound to act on a worst-case assumption and develop appropriate countermeasures. Such a course, some argue, could ultimately sap the effort by the Gorbachev government to cope with its internal economic problems, and perhaps even destabilize it.

Falling Ax

Federal outlays and commitments for research and development will drop precipitously this year as a result of the automatic budget-cutting procedure embedded in the Gramm-Rudman Act, which went into effect when Congress failed to meet its own target for reducing the national deficit. The reduction is 4.3 percent across the board for nonmilitary spending on R&D in the fiscal year ending September 30, and from 7 to 14 percent in commitments for future spending. Deeper cuts are likely on the military side if the Department of Defense exercises its option to protect operational strength and readiness at the expense of R&D and other programs.

The cutbacks reflect the impact of the law adopted by Congress last year in an effort to achieve a balanced budget by 1991. Under the law the deficit was to be reduced to $171.9 billion this year and is to be cut to $144 billion in fiscal 1987 and by $36 billion each year thereafter. Spending cuts go into effect when the normal budget process fails to meet these targets.

Congress limited the mandatory cuts to $11.7 billion for this year; since the projected 1986 deficit is $220.5 billion, the entire $11.7 billion must be excised now. The law requires a flat reduction of 4.3 percent in every nonmilitary "program, project and activity," with certain exceptions such as Social Security benefits and Medicare.

Willis H. Shapley of the American Association for the Advancement of Science has analyzed the effect of the cuts on R&D in terms of outlay (actual spending this year) and budget authority (commitments to support pro-

SCIENCE/SCOPE®

A new satellite provided the first telephone link between earthquake-stricken Mexico City and the Mexican consulate in Los Angeles, helping hundreds of anxious callers learn whether their relatives had survived the disaster. The consulate, located in the city with the largest population of Mexican citizens outside of Mexico City, was flooded with calls after the 7.8-magnitude quake on Sept. 19. For help, it turned to Hughes Aircraft Company, which had built the country's three-month-old Morelos communications satellite. Hughes engineers located a shipment of communications equipment en route to New York City and diverted it to the satellite ground station outside the Mexican capital. Meanwhile, an antenna at the Hughes ground station near Los Angeles was pointed at Morelos. To complete the phone line, the engineers established a microwave link between the ground station and company offices, then hooked into the local phone system to the consulate. The line was kept open 24 hours a day.

A new processing technique eliminates impurities in an optical fiber that has promising uses in the mid-infrared region of 1 to 5 micrometers. Zirconium fluoride glass fibers, which are typically prepared in an atmosphere of inert gases, contain defects that scatter light transmissions and preclude their use in long fiber links. Scientists at Hughes Research Laboratories, however, have prepared molten glass at 850°C using a novel reactive atmosphere process. This special process competely eliminates the chemical interaction with impurities, which yield light-scattering defects.

Although already advanced, North America's air defense system will be improved in the next few years to become even more vigilant in protecting U.S. and Canadian skies. The Joint Surveillance System, developed by Hughes for the U.S. Air Force, spans the continent from Alaska to Florida and Labrador to Hawaii. Already Hughes is developing a new computer, called the 5118MX, which has 3 million words of memory and will be at least three times faster than the current computer. Eventually, too, radar information from E-3A AWACS (Airborne Warning and Control System) aircraft will be fully integrated with JSS to expand coverage more than 200 miles beyond U.S. and Canadian borders.

A private, domestic satellite system will carry telecommunications throughout Japan beginning in early 1988. The system will be owned and operated by Japan Communications Satellite Company, Inc., a joint venture composed of Hughes Communications, Inc. (a Hughes subsidiary) and Japanese partners C. Itoh & Company, Ltd. and Mitsui & Company, Ltd. The joint venture firm has ordered two large, high-power satellites based on the new Hughes HS 393 spacecraft. These satellites will allow users to receive voice, television, and data transmissions through small, low-cost ground terminals. Each satellite will have 32 transponders, providing capability to transmit 32 channels of TV programming or a mix of TV and other communications. The satellites are scheduled for launch in December 1987 and April 1988. Services are expected to begin in February 1988.

Hughes Research Laboratories needs scientists for a spectrum of long-term sophisticated programs, including: artificial intelligence knowledge-based systems and computer vision; applications of focused ion beams; electron beam circuit testing; liquid-crystal materials and displays; nonlinear optics and phase conjugation; submicron microelectronics; plasma applications; computer architectures for image and signal processors; GaAs device and integrated circuit technology; optoelectronic devices; and materials and process technologies for high-speed infrared detection and optoelectronic applications. Send your resume to Professional Staffing, Hughes Research Laboratories, Dept. S2, 3011 Malibu Canyon Road, Malibu, CA 90265. Equal opportunity employer. U.S. citizenship required.

For more information write to: P.O. Box 45068, Dept. 79-13, Los Angeles, CA 90045-0068

HUGHES
AIRCRAFT COMPANY

Subsidiary of GM Hughes Electronics

The personal computer that raised high performance to new heights.

If you work with high volumes of information, you need answers fast.

You need a personal computer that's up to the task.

Which is why IBM created the Personal Computer AT® system. It's changed a lot of ideas about business computing.

The idea of "fast" has become much faster. The idea of "data capacity" has become far greater.

There are new definitions of "power" in a stand-alone PC. While phrases like "sharing files" and "multi-user systems" are being heard more often.

And surprisingly, words like "affordable" and "state-of-the-art" are being used *together.*

Clearly, the Personal Computer AT is different from anything that came before. And what sets it apart can be neatly summed up in two words.

Advanced Technology.

If you've ever used a personal computer before, you'll notice the advances right away.

To begin with, the Personal Computer AT is extraordinarily fast. That's something you'll appreciate every time you recalculate a spreadsheet. Or search through a data base.

It can store mountains of information—literally thousands of pages' worth—with a single "hard file" (fixed disk). And now you can customize your system to store up to 30,000 pages with the addition of a *second* hard file.

The Personal Computer AT runs many of the thousands of programs written for the IBM PC family. Like IBM's TopView, the program that lets you run and "window" several other programs at once.

Perhaps best of all, it works well with both the IBM PC and PC/XT. Which is welcome news if you've already made an investment in computers.

You can connect a Personal Computer AT to the IBM PC Network, to share files, printers and other peripherals with other IBM PCs.

You can also use a Personal Computer AT as the centerpiece of a three-user system, with your existing IBM PCs as workstations.

Most important, only the Personal Computer AT offers these capabilities *and* IBM's commitment to quality, service and support. (A combination that can't be cloned.)

If you'd like to learn more about the IBM Personal Computer AT, see your Authorized IBM PC Dealer, IBM Product Center or IBM marketing representative. For a store near you, call 1-800-447-4700 (in Alaska, call 1-800-447-0890).

The IBM Personal Computer AT, for Advanced Technology.

IBM, Personal Computer AT, PC/XT and TopView are trademarks of International Business Machines Corporation.

IBM®

THE SCIENCE OF WILDLIFE MANAGEMENT:

The Future of our Wildlife Depends on It

#1

Research

The precarious future that many species of North American wildlife faced around the turn of the century provided the impetus for the establishment of our first wildlife parks and refuges. Though initially effective, these early efforts aimed at helping wildlife soon developed serious shortcomings. The concept of providing complete protection, including the elimination of natural predators, to certain species was successful in building up threatened herds of animals, including elk and deer; however, as early as the 1920s, populations in many areas were outstripping their available food supplies.

Such problems helped spur the rapid growth of the modern science of wildlife management. Early wildlife management professionals were the first to recognize the vital importance of vegetation and other aspects of the natural environment that supported wild animal populations. This new understanding of the relationship between wildlife and habitat helped lead to the practical steps necessary to ensure the long-term abundance and health of certain kinds of wildlife.

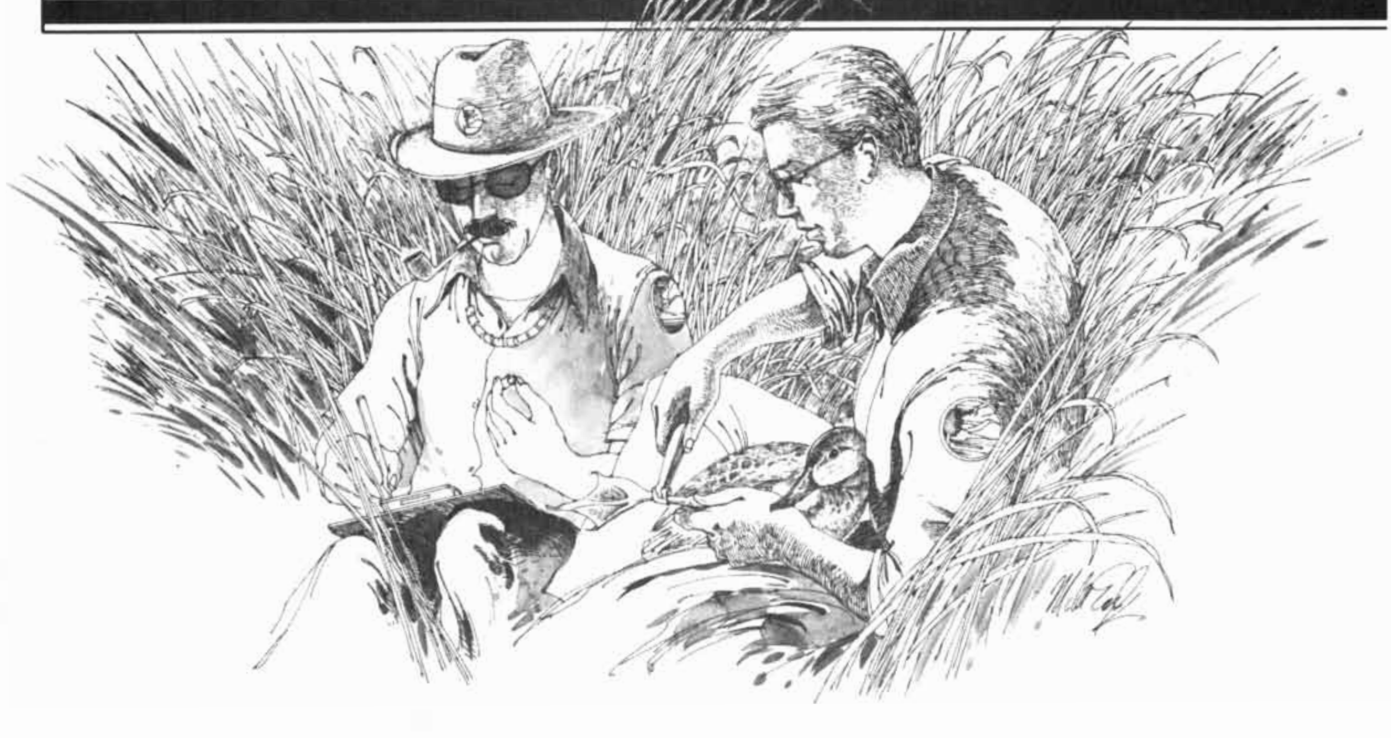
Extensive biological research is the foundation on which all management programs are built. Studies on animal numbers, their distribution, food preferences and the like provide a detailed picture of a species' needs and habits.

Bird banding projects, such as these, help determine a species' seasonal and local movements and can provide information on age, longevity and other vital characteristics important in developing successful conservation programs.

Over the years, it has been the American hunter who, through license fees and excise taxes, has provided the lion's share of the funds necessary for these conservation programs.

NSSF

National Shooting Sports Foundation
1075 Post Road
Riverside, Conn. 06878



grams, mostly by spending in future years). He estimates that outlay reductions will average 4.6 percent for each nonmilitary, nonexempted program and 3 percent for military programs. Budget-authority cuts will range from 7 to 14 percent for nonmilitary programs and from 4.5 to 9 percent for defense R&D.

In 1987, when the full amount needed to get the deficit down to $144 billion must be cut, the situation will be even more critical. "The prospects," Shapley says, "are for substantial, perhaps deep, reductions in currently approved levels of R&D funding."

Man Bites Dog

The rejection of a grant by a university is news for the same reason that a man biting a dog is news. And so Cornell University made news when it turned down a grant of $10 million from the Federal Government for a supercomputer. The refusal to accept the grant, university officials say, deliberately underscores the institution's opposition to the propensity of Congress for awarding backdoor grants, thereby short-circuiting the normal merit-review process. If such a practice becomes the norm, Cornell and other university officials fear, merit will cease to be a factor in the funding of research proposals: the pork barrel will replace peer review.

The impetus to seek funding outside the peer-review process has several sources. According to Robert Rosenzweig, president of the Association of American Universities, there is sometimes a sense of desperate need, arising from the fact that Federal funding for research has declined severely over the past 20 years. Sometimes Congress becomes a court of last resort when an institution believes it has been or will be slighted in the reviewing process.

The appropriation for Cornell has a very different history. A conference-committee report on the final version of a bill providing funds for the Department of Defense mandated the apportionment of $65.6 million in research grants among 10 universities. The provision included specific instructions to the Defense Advanced Research Projects Agency (DARPA) to buy a supercomputer from Floating Point Systems of Oregon for use at Cornell. The author of the provision was Senator Mark Hatfield, the Oregon Republican who is chairman of the Senate Appropriations Committee. None of the research grants had been through the merit-review process, and some of them had not even been reviewed by the congressional committees concerned.

When word of the $10-million grant reached the university, Frank Rhodes, the president, sent a "message to members of the Congress" that Cornell "will not accept funding awards which bypass normal review procedures." John F. Burness, vice-president for university relations, said: "The issue is what's in the best interest of the country and science—grants based on the merits of the project or grants awarded through political influence."

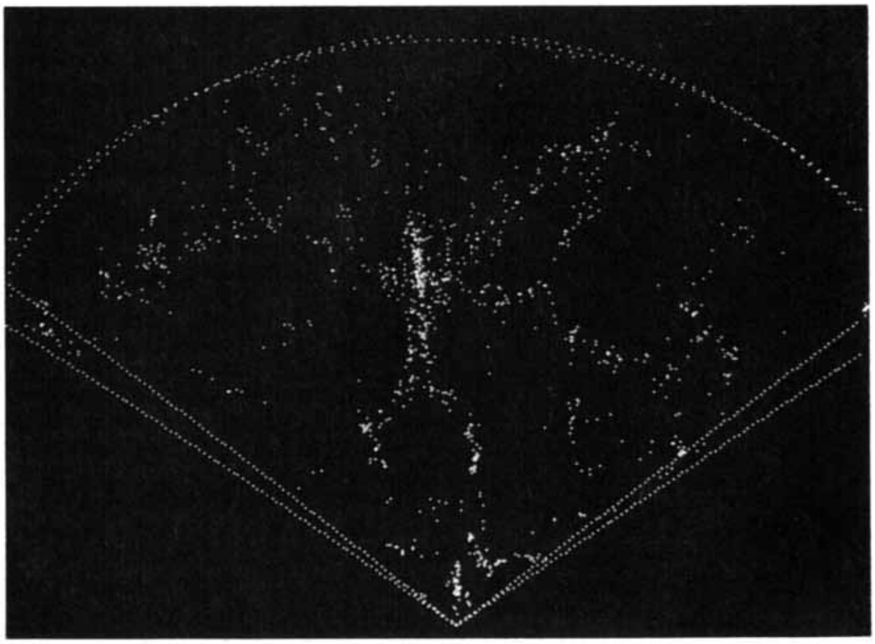
This time at least the price of principle was low. As it happens, Cornell and Floating Point have been working together to develop the supercomputer, and the university has submitted to DARPA a proposal under which the machine would come to Cornell. After the usual merit-review process it probably will, but in the right way from Cornell's point of view, and for the right reasons.

Cosmic Cartography

Astronomers from the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory have constructed a three-dimensional map of a wedge of the universe that is causing their colleagues to review ideas about the birth of the cosmos. Analysis of the map indicates that the universe is made up of gigantic bubbles: spherical or slightly elliptical regions of space apparently void of matter, whose outer surfaces are defined by galaxies. "If we're right, these bubbles fill the universe just like suds filling the kitchen sink," comments John P. Huchra, one of the investigators. Huchra and two of his collaborators, Valérie de Lapparent and Margaret J. Geller, note in *Astrophysical Journal Letters* that the immense size of the bubbles suggests that powerful stellar explosions—and not the force of gravity as is widely thought—had the primary role in the formation of the universe.

Working with a 1.5-meter telescope at the Fred Lawrence Whipple Observatory in Arizona, the investigators mapped more than 1,000 galaxies confined within a pie-shaped wedge of the sky measuring 120 degrees by six degrees. The map is based on an underpinning of modern astronomy known as Hubble's law. According to a theory proposed half a century ago by the American astronomer Edwin P. Hubble and subsequently verified by countless observations, the universe is expanding uniformly in all directions. Moreover, the velocity of any galaxy as seen from any other one is proportional to the distance between them. It turns out that measuring the velocity of a galaxy is fairly easy: light emitted by a receding galaxy is shifted toward the red end of the spectrum. By measuring the amount of the "red shift" of a galaxy one can therefore determine its recessional velocity and hence its distance from the observer's position in the universe.

All the galaxies on the map are 300



THREE-DIMENSIONAL MAP of a wedge of the universe shows galaxies (*represented by dots*) distributed on the surfaces of giant bubblelike structures. The apparent human torso at the center is a cluster of galaxies in the constellation Coma. The map was prepared by Margaret J. Geller, John P. Huchra and Valérie de Lapparent of the Center for Astrophysics of the Harvard College Observatory and Smithsonian Astrophysical Observatory.

GM
LETS GET IT TOGETHER....BUCKLE UP.

million light-years or less from the earth. They lie on the surfaces of bubbles that measure from about 60 to 150 million light-years across. (Our own galaxy, the Milky Way, measures about 60,000 light-years across.) Although astronomers had earlier identified one of the largest such bubbles, they had believed it was an anomaly. "It now turns out that this type of structure is very common, but we couldn't see it before because we hadn't looked at a large enough volume of the universe," comments Geller. "We probably are sitting on the surface of a bubble. We can't see the other bubbles until we look deep enough."

Prevailing theories of the history of the universe, particularly the so-called bottom-up and top-down (or pancake) theories, seem unable to account for the existence of the bubbles. According to the bottom-up theory, galaxies formed out of a gaseous soup; the galaxies subsequently coalesced to form clusters and superclusters. The top-down theory, on the other hand, holds that matter in the universe initially collapsed into vast pancakelike sheets; the pancakes then fragmented, giving rise to galaxies, clusters and superclusters. Neither model predicts the formation of bubbles that have the sharply defined surfaces seen in the Harvard-Smithsonian survey.

The bubble theory goes back to 1981. It was then that Jeremiah P. Ostriker of Princeton University and Lennox L. Cowie of the Space Telescope Science Institute and Johns Hopkins University proposed that when the universe was about a billion years old, massive stars collapsed and exploded in supernovas. (The universe is widely held to be from 15 to 20 billion years old.) The explosions created shock waves that drove matter into spherical shells. Although the shells predicted by the Ostriker-Cowie model are smaller than the observed bubbles, that discrepancy could be accounted for in several ways. Ostriker notes, for instance, that the largest bubbles may be consolidations of smaller ones.

During the next few years the Harvard-Smithsonian investigators would like to increase the area covered by their map by a factor of 10. They would also like to determine whether the bubbles are really as empty as they appear to be.

Tour de Force

In modern physics no more than four fundamental forces have sufficed to explain the nature of everything in the universe—from the behavior of subatomic particles to that of the largest galactic supercluster. Recent analysis of data from three disparate experiments, including one that was carried out at the beginning of the century, suggests that a complete picture of physical reality may require the existence of a fifth force.

Ephraim Fischbach, currently at the University of Washington in Seattle, Daniel Sudarsky, Aaron Szafer and Carrick Talmadge of Purdue University and Samuel H. Aronson of the Brookhaven National Laboratory have proposed in *Physical Review Letters* the existence of an intermediate-range force to complement the gravitational and electromagnetic forces, which are effectively infinite in range, and the so-called strong and weak nuclear forces, which do not extend beyond the radius of an atomic nucleus. Their suggestion is based on investigations that were originally spurred by the anomalous behavior of the subatomic particles called *K* mesons in high-energy experiments done about 10 years ago. The group was further intrigued by measurements appearing to show that gravity operates differently on objects in deep Australian mine shafts than it does on objects at the earth's surface.

Could the puzzling discrepancies in both experiments be related? The investigators thought they could. They postulated that a weak, repulsive force whose range is several hundred meters could account for the observed results. The *K*-meson experimental results suggested such a force would have to depend on hypercharge: a fundamental attribute, like charge or spin, that is assigned in integer units to the more massive subatomic particles such as *K* mesons or nucleons (the proton and the neutron).

To test their hypothesis the group looked for other published results of experiments sensitive enough to reveal the subtle effect of a putative fifth force. One promising candidate was the set of experiments, carried out between 1889 and 1908, in which the Hungarian physicist Roland von Eötvös compared, with an accuracy of one part in a billion, the gravitational acceleration of various materials toward the earth.

In examining the data from the Eötvös experiment the investigators report they did indeed find evidence of a force whose parameters fitted closely with those they postulated, including the dependence on hypercharge. Eötvös had dismissed small inconsistencies in his data as being statistically insignificant. The workers found a systematic effect among them, however: the force exerted by the earth on various materials appeared to vary slightly with the number of nucleons per unit of mass (which varies from element to element, and indeed from isotope to isotope).

The reexamination of the 80-year-old data has caused some furor because the Eötvös experiment has been held to provide conclusive evidence that the gross mass of an object, and not its specific composition, determines the force between it and the earth, which is to say its weight. The authors of the study stress the tentativeness of their analysis and emphasize that further experimentation is necessary before any definitive statement can be made about a new, hypercharge-dependent force. They suggest several possible tests. For example, the Eötvös experiment could be repeated with greater sensitivity. Or one might go 300 years further back in time to update, with modern instrumentation, the classic experiment in which Galileo allegedly dropped test masses of differing composition from the same height and measured the time of flight.

Expert on a Chip

An integrated-circuit chip has been made that can render so-called expert decisions with great speed on the basis of imprecise information. The device is designed for controlling automated machinery that must function under constantly changing and uncertain circumstances.

Expert systems have been around for a number of years in manufacturing processes and to some extent in medicine, embodied as specialized software for general-purpose computers. By distilling the knowledge of human experts into a set of simple, interlinked rules, software engineers were able to write computer programs that simulate an expert's reasoning. Such programs typically accept data and sift through the network of rules until a chain of rules linking the given input conditions to an appropriate output is found.

Two major problems hamper such programs. Retrieval of the program's instructions and decision-making rules from an external memory requires a certain amount of time; such a system may not be suitable for "real time" applications in which input data keep changing. The other problem is that the data often require precise definition; if the input does not perfectly match situations described in the rules, many expert systems cannot respond.

The expert-system chip designed and built by Masaki Togai and Hiroyuki Watanabe at the AT&T Bell Laboratories also relies on a collection of rules to come to a decision. But because its operating instructions are wired directly into the microcircuitry, the chip can function about 10,000 times faster than conventional expert systems. Moreover, the chip makes use of what is called fuzzy logic, with the result that input conditions do not have to be accurately specified and rules do not have to be expressly written for every possible situation. In the chip's logic "maybe" is just as legitimate an input as "yes" or "no."

The built-in fuzzy logic enables the chip to accept information that is not well defined and to compare it to all the stored rules simultaneously, assigning a weight to each rule according to how well each matches the data. The final decision is then based on the combined recommendations of the rules with the highest weights. As a result the chip reasons with human flexibility—but at superhuman speeds.

Destructive Sway

Off the west coast of Mexico the Cocos plate, a moving piece of the earth's outer shell, dives eastward under the North American plate, carrying part of the Pacific floor with it. On the average the plates grind past each other at a rate of six centimeters per year. As they do so large strains develop at points of high friction on the fault surface between them. Last September 19, at 7:18 in the morning, roughly 15 kilometers under the coastal town of Lazaro Cardenas, one of those points ruptured. Within seconds the rock masses on opposite sides of the fault were offset by about two meters. Within a minute the rupture spread 170 kilometers along the fault. Although the resulting tremors registered 8.1 on the Richter scale, damage within the rupture zone was comparatively slight. But 360 kilometers away, in Mexico City, violent ground motion damaged hundreds of buildings and claimed thousands of lives. The tragedy is now being attributed to unpredictable geologic and architectural coincidences.

The earthquake itself was not unexpected, because it occurred on a stretch of the plate boundary where there had been no large rupture in 74 years. In anticipation of a quake John G. Anderson, James N. Brune and their colleagues at the University of California at San Diego, along with Jorge Prince and his co-workers at the National Autonomous University of Mexico, had deployed an array of strong-motion accelerographs in the area. The peak ground acceleration near the epicenter on September 19 was only 15 percent of gravitational acceleration, which is surprisingly mild: accelerations as powerful as 1 *g* have been recorded during much smaller quakes in California.

The strongest ground motion was recorded not near the fault but in the center of Mexico City. The reason seems to be that most of the city was built on a layer of clay left by an ancient lake. Like any physical system, the clay layer resonates when it is stimulated by vibrations of the right period. The resonance period can be calculated from the thickness and seismic velocity of the layer; according to James L. Beck of the California Institute of Technology, it turns out to be about two seconds. As a result seismic waves with a two-second period are amplified dramatically in the clay. Whereas the ground motion observed in most areas after an earthquake displays a broad spectrum of periods, with most of the energy carried by waves with periods well below two seconds, in Mexico City last September the ground motion was dominated by two-second waves.

The period of the motion is just as important as its amplitude in determining how much damage it causes, because buildings too can resonate. A building's period can be estimated by allowing a tenth of a second per floor. By this rule of thumb buildings of about 20 stories should have sustained the most damage in Mexico City, but they did not. Most of the 300 or so damaged buildings had from six to 15 stories. According to Beck, the reason was that a building's resonance period increases as it is shaken; internal partitions and masonry give way and the building is in effect softened. Hence the middle-size buildings in Mexico City were pulled into resonance with the two-second ground motions, while taller buildings moved out of resonance. Ironically, most of the low buildings built before the introduction of a modern building code in 1957 remained intact.

The authors of the building code were aware of the implications of the clay layer; the design requirements in Mexico City for buildings with long periods are more stringent than they are in the U.S. What the code drafters could not have foreseen, says Beck, is the extraordinary duration of the September earthquake. Teleseismic records analyzed at Caltech by Holly Eissler, Luciana Astiz and Hiroo Kanamori suggest that the quake actually consisted of two distinct shocks about 25 seconds apart, each about 16 seconds long. The violent shaking in Mexico City lasted for more than a minute. With each resonant cycle the amplitude of the motion in the clay layer under the city increased, and more buildings were brought under its destructive sway.

Unequal Work, Unequal Pay

During the 1970's women made significant gains in entering occupations once dominated by men. Yet those gains have by no means eliminated the problem of sex segregation in the work force. Most men and women in the U.S. continue to work in jobs that are segregated by sex, resulting in lower pay for women. Furthermore, if the Federal Government abandons its role as an advocate of occupational integration, progress in the next decade could be very slow.

Those are some of the conclusions of a report issued recently by the National Research Council of the National Academy of Sciences. According to the report, in 1980 some 48 percent of all U.S. women worked in occupations that were at least 80 percent female. Typical female occupations included those of secretary (99 percent female) and registered nurse (96 percent female). Among men segregation was even more extreme: 71 percent of men worked in occupations that were 80 percent or more male. Male occupations included those of carpenter (only 1 percent female) and engineer (4 percent female).

The division between men's work and women's is reinforced at the level of the individual company. The report cites a recent study of sex segregation among businesses in California. The study found that of 393 firms, 30 hired workers of one sex only. In 201 additional businesses men and women shared no job titles. Only a small minority of companies appeared to be relatively integrated according to sex. Even there the appearance was deceptive, because men and women worked side by side at the same job in only a few instances.

Segregation by occupation is largely responsible for the significant difference in income between men and women. In 1981 the median annual income for men was $20,260; for women it was $12,001, or about 60 percent of men's earnings. If there had been no occupational segregation, women would have earned 75 percent of what men earned, the report concludes. If segregation within job categories at individual companies were also eliminated, much of the remaining 25 percent discrepancy would disappear.

Although such statistics seem discouraging, women made more progress in the 1970's than they had in the preceding decades. From 1900 to 1970 the index of occupational segregation remained in the high 60's. (The index represents the proportion of male and female workers who would have to move to fields dominated by the opposite sex for occupational segregation to be completely eliminated.) Between 1970 and 1980 the index underwent the largest single decline of the century—almost 10 points—and reached a level of about 60.

Progress in the coming years may be slow, however. According to the report, Federal intervention has been one of the most effective methods of overcoming occupational segregation. Employers have made voluntary efforts to reduce segregation after observing the results of sex-discrimination lawsuits brought by Federal authorities. The authors of the report conclude that the recent decrease in Federal efforts to enforce civil-rights legislation is likely to impede progress, largely because employers, no longer afraid of Federal sanctions, will fall back on discriminatory habits.

Costs of Commercialization

Hospitals affiliated with investor-owned chains reap the benefits of centralized, high-caliber management and economies of scale, making them more efficient than their not-for-profit counterparts. So say many who favor investor ownership of hospitals, but a study published in the *New England Journal of Medicine* suggests otherwise. The authors, J. Michael Watt, Robert A. Derzon and James S. Hahn of Lewin and Associates, Inc., and Steven C. Renn, Carl J. Schramm and George D. Pillari of the Johns Hopkins School of Public Health, report that the investor-owned hospitals whose financial performance they examined incurred about the same costs per case as matched not-for-profit hospitals did. The investor-owned institutions earned higher profits, but only because they charged more for comparable services.

The sample included 80 general hospitals belonging to investor-owned chains in eight southern and western states; each hospital was paired with a not-for-profit hospital that has similar services and a comparable size, location and average length of stay. The authors also tried to control for differences in the quality of care. For each matched pair the authors reviewed financial data for the fiscal years 1978 and 1980.

The study revealed little difference in the amounts the two kinds of hospitals spent on most categories of inpatient services other than drugs and medical supplies, which cost investor-owned hospitals about $7 a day more for each patient than they cost the not-for-profits. The investor-owned institutions did spend less to operate and maintain their facilities. Those savings were more than offset by the expense of maintaining a corporate headquarters and the higher plant-depreciation costs of the investor-owned hospitals, which often have newer facilities than not-for-profit hospitals.

In spite of their comparable efficiency and the fact that investor-owned hospitals must pay taxes and do not receive the charitable contributions and public funds available to some not-for-profit institutions, the investor-owned group earned higher net incomes and profits. They did so, the investigators found, by charging a total of about 22 percent, or $401, more than the not-for-profit hospitals during fiscal 1980 for each patient admitted. Higher charges for ancillary services, such as laboratory services, drugs and oxygen therapy, accounted for most of the price difference; routine charges, such as the basic room rate, were only slightly higher at the investor-owned institutions. "These results," the authors write, "suggest the existence of a strategy by the investor-owned chain hospitals of setting competitive prices for the more visible 'room and board' services while setting higher prices for ancillary services, which are less easy to compare from hospital to hospital."

The authors suggest that the investor-owned chains' practice of maximizing profits by charging more rather than by pursuing operating efficiencies is not surprising in view of the traditional policies under which Medicare, Medicaid and private insurers have reimbursed hospitals either for the cost of providing services or for all billed charges. Since the period covered by the study many of those policies have been revised, however, so that both categories of hospitals will have to adapt to cost-containment measures. For the investor-owned hospitals, the authors write, "the major challenge of the new cost-conscious environment will be the achievement of economies of scale that theory suggests should be possible through system operation."

Enhanced Promotion

All mammalian cells contain the gene coding for insulin, and yet insulin is produced only by the beta cells of the pancreas. How is it that a differentiated cell in a multicellular organism expresses only a particular fraction of its genetic inheritance? Some important pieces of the puzzle have begun to come clear.

One line of investigation has focused on uncovering the mechanisms that activate the first step in gene expression: the transcription, or copying, of DNA into RNA for subsequent translation into the protein the gene encodes. As in bacteria, transcription in eukaryotic cells (the nucleated cells of higher organisms) is known to begin when the enzyme RNA polymerase, which mediates transcription, binds to promoter sequences adjacent to the site where transcription begins. Eukaryotic cells are subject to additional transcription controls, including what are known as enhancer sequences. These elements, originally observed in mammalian viruses, increase the activity of nearby promoters and the rate of transcription, possibly by facilitating the binding of RNA polymerase to the promoter. Cellular enhancers identified thus far have been found to be cell-specific: they function only in a given cell type, presumably by interaction with proteins present only in that cell. The putative activating proteins for cellular enhancers are still being sought.

Among the most recently discovered cell-specific enhancers is the one for the rat insulin gene, which was identified by William J. Rutter, Michael D. Walker and their colleagues at the University of California at San Francisco together with Thomas Edlund of the University of Umeå in Sweden. The group, which reported its findings in *Science*, identified the enhancer as a region of about 200 base pairs (the runglike subunits of the DNA double helix) immediately upstream from (to the left of) the promoter. They did so by systematically deleting segments of the putative enhancer region and noting which were needed for transcription in beta cells.

The group went on to show that this enhancer region is selectively activated in beta cells. DNA containing the enhancer was attached to a "promiscuous" viral promoter and a bacterial gene coding for a bacterial protein. When the recombinant DNA was inserted into various cell lines, it stimulated high levels of bacterial-gene expression only in beta cells. Since the chosen promoter would have been active in any cell type, the investigators concluded that the increased expression in the beta cell must be caused by the only potentially selective element: the enhancer. Then they showed that the insulin gene's promoter is also preferentially activated in beta cells. A bacterial gene fused to an insulin promoter and a promiscuous enhancer was expressed 10 times as much in beta cells as in other cells. The investigators suspect that the combined selective effects of the enhancer and promoter may be greater than the sum of the parts, which could explain the very

SEEING THE LIGHT

A special selection of exhibits from San Francisco's Exploratorium

IBM GALLERY OF SCIENCE AND ART

STOP

TIME

SEE

SOUND

DISTORT

SPACE

DEFY

GRAVITY

MIX

WAVES

BEND

LIGHT

COLOR

SHADOWS

SPLIT

IMAGES

ENTER

INFINITY

Madison Avenue at 56th Street · January 29th-April 26th, 1986 · Tues. to Fri., 11 am-6 pm · Sat., 10 am-5 pm · Free admission

IBM

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.

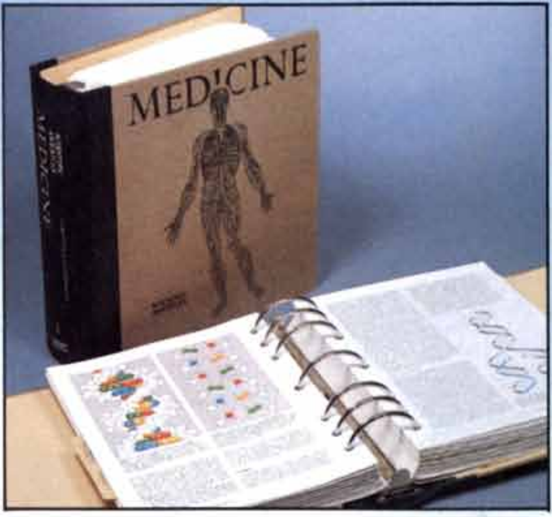
Review courses cover clinical advances—but, months later, do you recall the details of a new procedure or unfamiliar drug?

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

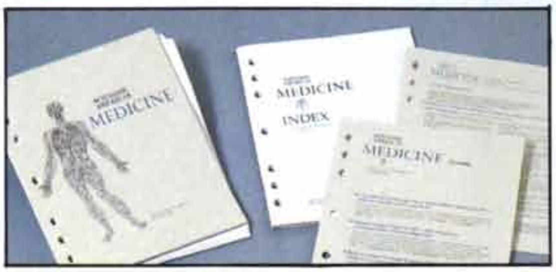
MEDI

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 ____________________

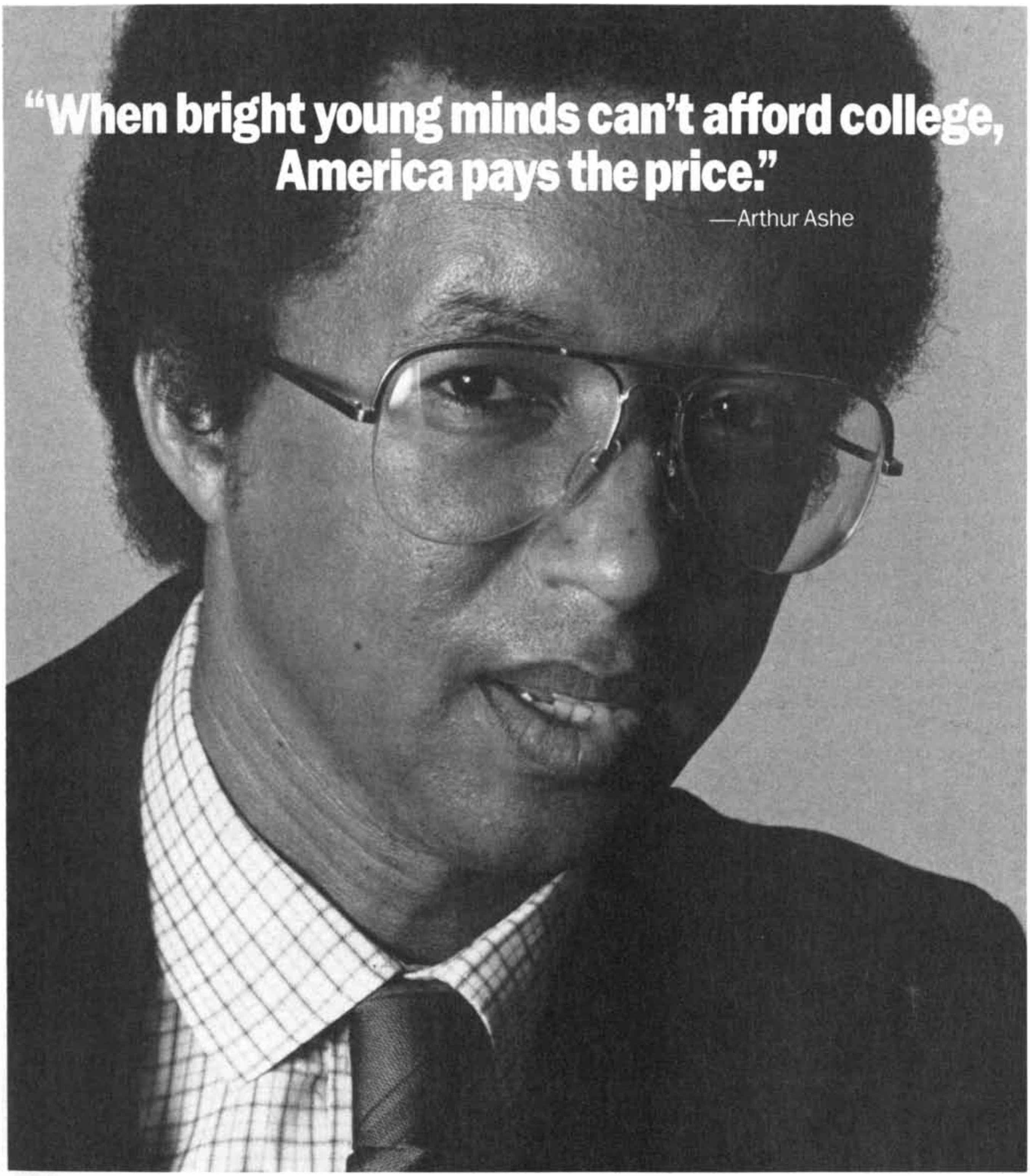
Specialty ____________________

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.

7C


"When bright young minds can't afford college, America pays the price."
—Arthur Ashe

high insulin output seen in pancreatic beta cells and, indeed, the high productivity of selectively expressed genes in other highly specialized cells.

Rutter's group and others believe isolation of enhancers will allow workers to identify and isolate the activating proteins that are specific for each cell type (which Rutter has named differentiators). Such discoveries should then lead to the identification of the genes encoding those unique proteins and ultimately, perhaps, to "master control" genes in each cell type that switch on the rest of a cell's controls.

Eventually it might be possible to coax selected cells into carrying out new functions. For instance, in insulin-dependent diabetes the cells producing insulin are destroyed. The disease might be treated by activating the silent insulin genes in other hormone-secreting cells and causing them to produce insulin inside the body.

A Case of Nerves

Most of the considerable attention that has been devoted to AIDS has focused on the capacity of the causative virus to weaken the immune system and render the victim fatally vulnerable to a host of other disorders. Now it appears that the virus, human *T*-cell lymphotropic virus III (HTLV-III), is also capable of causing disease directly. Two reports in the *New England Journal of Medicine* show that the virus can infect cells of the nervous system, causing meningitis and degenerative changes in brain tissues.

The two studies may herald a significant change in how clinical investigators regard the AIDS virus. Since its discovery the virus has been known to have an affinity for some of the white blood cells known as lymphocytes. Indeed, it is by subverting the lymphocytes called helper *T* cells (which have a critical role in triggering an immune response) that the pathogen undermines the body's defenses against disease. In the light of the recent findings it would seem the virus also has a strong affinity for cells of the central nervous system.

Such affinity helps to explain the meningitis and diffuse degenerative changes in the brain that are common among AIDS victims. One *New England Journal* article reports the work of a group that cultured HTLV-III from brain tissues, spinal-cord tissues and cerebrospinal fluid of AIDS victims who displayed neurological symptoms. Samples from seven of nine patients with meningitis and 10 of 16 with dementia (resulting from brain degeneration) yielded the virus. Those findings are reinforced by the work of the second group, which found antibodies to HTLV-III in the cerebrospinal fluid of AIDS patients who had neurological disease.

In order to reach the cerebrospinal fluid the virus must somehow cross the blood-brain barrier: the dense physical and physiological barrier that separates the blood, and much of what it carries, from the cells of the central nervous system. How the passage is achieved is not yet known, according to David D. Ho of the Harvard Medical School, leader of the group that cultured the virus from brain tissues. According to Ho, it seems likely that HTLV-III first enters cells in the blood. Some blood cells are capable of crossing the blood-brain barrier, and the virus may exploit them as vehicles to enter the brain.

The clinical effects of the virus in its passage from the bloodstream to the brain can vary greatly. Ho noted that some patients in his group's study had neurological symptoms but no immunologic deficiency; others showed immunologic deficits but no brain symptoms; still others had both types of disorder. How the AIDS agent gives rise to such disparate clinical results is not yet known.

The discovery that HTLV-III can infect brain tissue has disturbing implications for potential AIDS therapies. Several viruses are known to be able to enter nerve cells and remain latent there without producing symptoms, erupting periodically to cause episodes of disease. HTLV-III is closely related to at least one such pathogen, the visna virus, which causes a degenerative neurological disease in sheep. If the AIDS virus can retreat into the nervous system, it might be extremely difficult to eradicate, because the blood-brain barrier is impenetrable to many therapeutic substances.

Demon Rum

Ethyl alcohol, or ethanol, the active constituent of wine, beer and liquor (and thus the world's preeminent mood-altering drug), is earning a place among the knottier mysteries confronting the neurosciences. A recent finding points up the difficulty of discovering precisely what the drug does in the brain to bring on its effects.

At the Scripps Clinic and Research Foundation in La Jolla, Jorge R. Mancillas, George R. Siggins and Floyd E. Bloom investigated (in rats) the action of ethanol on neurons in the hippocampus, a part of the cerebral cortex. Earlier work at Scripps had established that the stimulation of neural signal paths leading to the hippocampal neurons called pyramidal cells causes the cells to increase and then curtail their electrical activity: the cells are excited and then inhibited. Remarkably, the presence of ethanol in the hippocampus had proved to amplify both phases of the pattern.

Mancillas, Siggins and Bloom undertook to discover how this dual amplification comes about. Several neurotransmitters, or brain messenger substances, are known to act in the hippocampus; some are excitatory, others inhibitory. By means of micropipettes the experimenters introduced small amounts of each neurotransmitter into the vicinity of hippocampal pyramidal cells, thus mimicking the release of neurotransmitter that constitutes the dispatching of a signal from one neuron to another. They monitored the electrical activity of the cells. When ethanol was introduced into the rats' bloodstream at concentrations sufficient to bring on intoxications ranging from mild to heavy, the transmitter acetylcholine proved to gain in its ability to excite the pyramidal cells; somatostatin gained in its ability to inhibit the cells. The changes were evident some 10 or 15 minutes after the ethanol entered the animals' blood and faded within an hour or two. Other neurotransmitters showed no change in potency in the presence of ethanol.

The hippocampus is known to be involved in the ability to commit things to memory; thus the new finding may ultimately have a place in explaining the impairment of memory characteristic not only of alcoholics but also, temporarily, of social drinkers. The hippocampus, moreover, is part of the limbic system, a set of brain structures that are known to influence mood. Still, the investigators, who report their finding in *Science,* "prefer not to generalize from observations of specific actions of ethanol on a given neurotransmitter, brain region, or neuronal type." Ethanol's "primary target" in the complexities of neuronal function remains to be named.

Why is the action of ethanol proving so hard to identify? Siggins offers several reasons. Ethanol is notably weak: almost all other psychoactive drugs are effective at far lesser concentrations. Ethanol seems not to have receptors: specific binding sites on the surface membrane of certain neurons. It simply enters every neuron by diffusing through the membrane. The action of ethanol depends markedly on the dose and on the time course of the administration of the drug. The action varies from neuron to neuron, and even among parts of individual neurons. Siggins is not surprised that "it's taking brain scientists so long to find out what's going on."

The Superconducting Supercollider

An accelerator 20 times as powerful as any now operating could be built by 1995. It would probe matter in unprecedented detail and re-create conditions prevailing near the beginning of time

by J. David Jackson, Maury Tigner and Stanley Wojcicki

The year is 1995. A pastoral landscape of farmland or prairie gives almost no hint that a tunnel, large enough to walk through and curved into a ring some 52 miles around, lies buried below the surface. Inside the tunnel there is a small tramway for maintaining two cryogenic pipelines, each about two feet in diameter. Within each pipeline is a much smaller, evacuated tube that carries a beam of protons, which are kept on course by powerful superconducting magnets surrounding the tube. With every circuit of the ring the energy of the protons in the two beam pipes is boosted by a pulse of radio waves; in 15 minutes the protons are accelerated around the ring in opposite directions more than three million times.

Suddenly electromagnetic gates are opened and the beam paths are made to cross. Pairs of protons collide, and some of the energy of the collision can be transferred at a rate that far exceeds the instantaneous output of all the power plants on the earth into a region whose diameter is 100,000 times smaller than the diameter of a proton. There, for a time so brief that it is to the second what the second is to 100,000 times the age of the universe, we shall have a glimpse of the universe at the moment of creation. The energy that will be concentrated in that region at that instant is now found only in the rarest of cosmic rays, but such energy concentration was the prevailing state of the universe in the first 10^{-16} second after the big bang. New elementary particles that could materialize from the energy may show how to explain the origin of mass.

It is remarkable that such a vision is well within the reach of 20th-century technology. The tunnel, the pipeline, the attendant operating systems and an initial complement of particle detectors and computers can be built entirely with available technology—albeit on a scale never before attempted—at a cost of about $4 billion in constant 1986 dollars. Indeed, the basic design of the instrument is already being tested at an energy scale of about 1/20: the scale model is the particle accelerator known as the Tevatron, at the Fermi National Accelerator Laboratory (Fermilab). The scaled-up version would make it possible to study energetic processes that are not accessible to any other accelerator now in operation or seriously contemplated anywhere in the world.

The proposed machine builds on the accumulated experience of more than 50 years in designing and constructing particle accelerators. Moreover, since 1982 the community of high-energy physicists has devoted considerable thought and energy to the design of an economically realistic machine that will have the greatest chance of resolving outstanding theoretical questions about the ultimate constituents of matter. Because it is designed as a colliding-beam accelerator and because superconducting magnets play an essential role in minimizing its consumption of power, the instrument we are planning is called the Superconducting Supercollider, or SSC.

Energy and Luminosity

Particle accelerators can set off the most highly energetic reactions one can study under controlled conditions. Because mass and energy are equivalent, the maximum reaction energy fixes the maximum observable mass of the basic material entities that can be created in the laboratory. Hence the design of a particle accelerator determines the limits of direct, experimental knowledge about the fundamental structure of matter. The maximum observable energy and mass in a particle accelerator depend on the energy of the accelerated particle beams, the way the energy of the beams is released and the intensity of the beams.

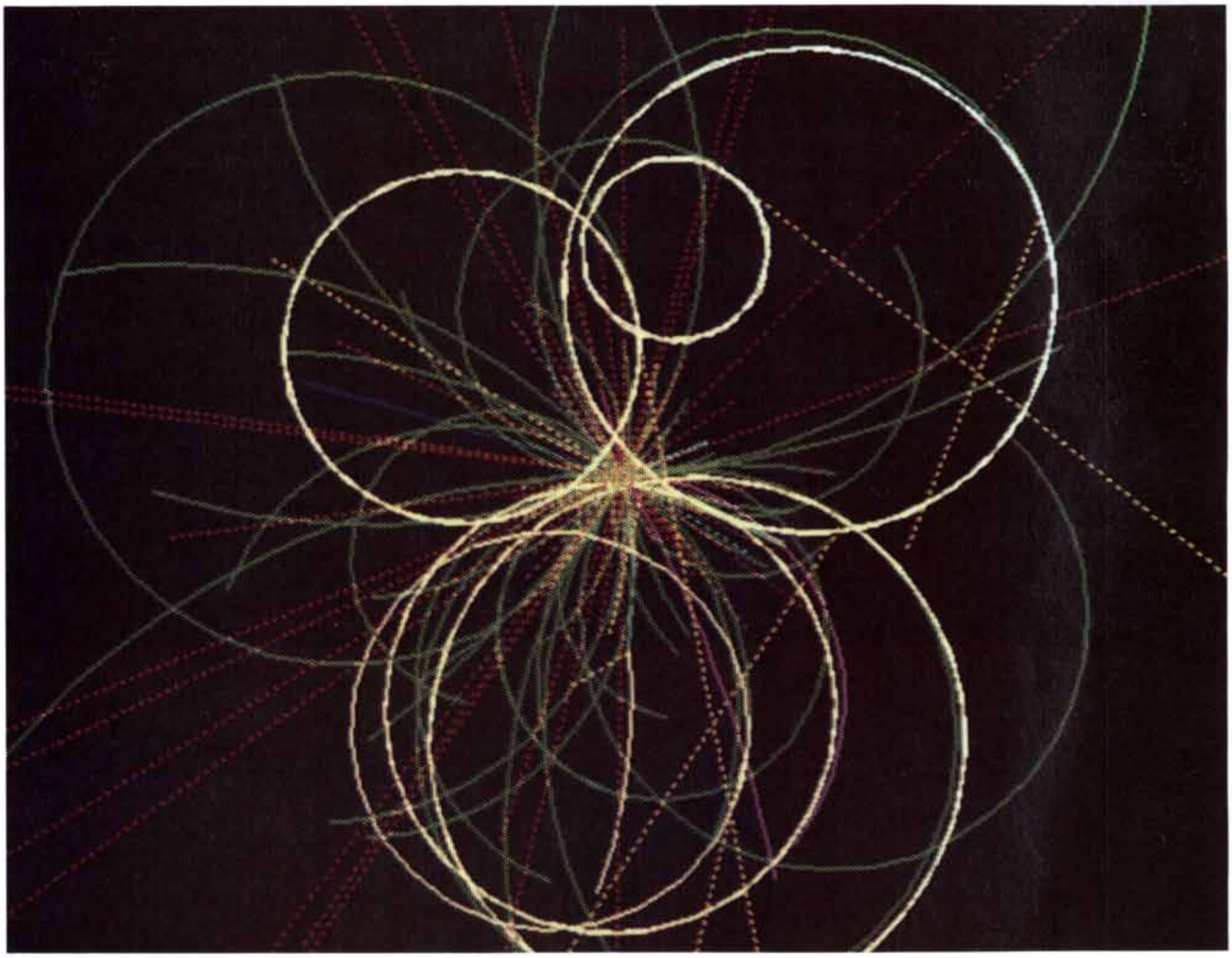
Energy can be released most effectively for creating new particles of high mass if two beams of equal and opposite momentum are made to collide. This conclusion follows from the law of conservation of momentum, which is strictly obeyed in all collisions between particles: the total momentum of the products of a reaction must be equal to the total momentum of the reacting particles. If a particle in an energetic beam is made to collide with a particle in a target at rest, the forward momentum of the energetic incoming particle must be conserved. Thus much of the total energy of the two particles is needed for imparting forward momentum to the particles leaving the scene of the collision.

In contrast, if the momentums of two colliding particles are equal in magnitute but opposite in direction, the total momentum of the pair is zero. In principle none of the total energy of the pair is needed for imparting momentum to the reaction products, and so much of that energy becomes available for creating new particles. Our design now calls for a machine in which two counterrotating beams of protons are each accelerated to an energy of 20 trillion electron volts (TeV) and made to collide; the energy of a collision would be 40 TeV, or more than 20 times the collision energy at the Tevatron collider. The energy available for creating new matter at the SSC will be 200 times the energy that would be available if only one such beam were directed at a fixed target.

In a colliding-beam accelerator a good measure of the useful beam intensity is the luminosity of the collider. Because the beams are made up of evenly spaced trains of equal bunches of particles, the luminosity is the number of particles per bunch in one beam multiplied by the number of intercepted particles per unit area in the second beam, all multiplied by the frequency with which the bunches collide. Luminosity is thus an index of the average number of events per second for any given reaction. A high luminosity is needed to guarantee that extremely rare high-energy reactions are generated often enough for physicists to confirm their existence. The current design of the SSC calls for a luminosity of 10^{33} in standard units, about 1,000 times that projected for the Tevatron.

Throughout the development of this design we have been acutely aware of the need for economic optimization; indeed, the SSC was recommended in 1983 by the High Energy Physics Advisory Panel to the U.S. Department of Energy only after an agonizing decision to abandon several other funding requests for particle accelerators. In response to the recommendation the Energy Department commissioned a research-and-development program, whose purpose is to prepare a technically realistic and economically optimum design and to give a detailed cost estimate. The program, coordinated by the SSC Central Design Group of the Universities Research Association, is being carried out at the Brookhaven National Laboratory, Fermilab, the Lawrence Berkeley Laboratory and the Texas Accelerator Center, in universities and in industry, with valuable assistance from abroad. We and our colleagues shall submit the report of that program next month, and it will serve as one important factor in a decision on further support of the SSC by the Energy Department.

Continued engineering development of the SSC will further lower its cost. Strive as we will, however, the final cost will be high compared with the costs of other scientific instruments. That cost is driven primarily by one of the basic laws of nature: in order to probe the structure of matter at increasingly fine resolution one must accelerate particles to increasingly high collision energies. To justify the costs of such a probe we shall try to explain why particle physicists think the extreme physical conditions available at the SSC are important to explore and



SIMULATED COLLISION of two protons, each accelerated to an energy of 20 trillion electron volts (TeV), gives rise to the shower of particles shown in the computer-generated image. Such energies exceed by a factor of 20 the energy that can be attained by the most powerful particle accelerators now in operation; they could be reached by the proposed accelerator, the Superconducting Supercollider (SSC). In the computer simulation a new particle called the Higgs boson, whose existence has been predicted by theory but has not yet been confirmed in an experiment, materializes out of the energy liberated at the site of the collision between the protons. The Higgs boson in the simulation has a mass of 300 billion electron volts (GeV), and it decays into a W^+ and a W^- boson, which mediate the weak force responsible for beta decay. Neither the Higgs boson nor the W^+ and W^- bosons live long enough to be detected, and so their existence must be inferred from their decay products. The W^+ boson decays into a positron (*solid blue line*) and a neutrino (*broken yellow line directed from center to upper left*), and the W^- boson decays into two light quarks, each of which gives rise to a shower of composite particles that is made up mostly of pions (*solid green lines*). Various other collision by-products are encoded by other colors: red for photons, white for muons and purple for baryons. The simulation was done by James Freeman of Fermi National Accelerator Laboratory (Fermilab), using the ISAJET model devised by Frank E. Paige, Jr., of Brookhaven National Laboratory.

how such conditions are made possible by the current design.

The Standard Model

In the past two decades remarkable progress has been made in identifying the basic constituents of matter and the fundamental forces by which they interact. According to what is now called the standard model of elementary processes, all matter is made up of quarks and leptons, whose interactions with one another are mediated by the exchange of so-called gauge particles. It is also thought there are four basic kinds of interaction: electromagnetic, weak, strong and gravitational.

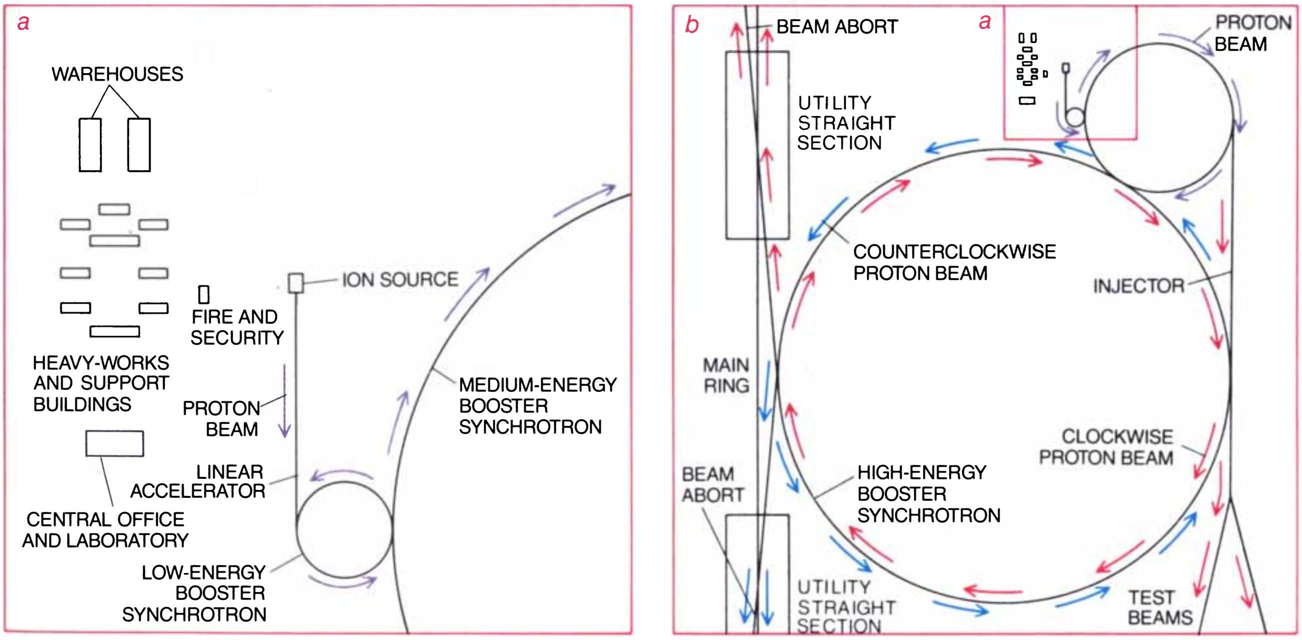
For example, the electron is classified as a lepton, and its electromagnetic interactions with the proton are mediated by a gauge particle called the photon. Beta decay, which is central to nuclear burning in the sun, is a result of the weak interaction, and it is mediated by the exchange of the gauge particles called weak vector bosons. The proton, the neutron and many other particles are classified as hadrons, and they are made up of three fractionally charged quarks. The quarks are held together by a strong interaction called the color interaction, and that interaction is mediated by the exchange of eight kinds of gauge particles called gluons. By analogy with these three interactions it is assumed that another gauge particle, the graviton, mediates the gravitational interaction, but such a particle has not been detected. In all it is now believed there are at least six quarks and their six corresponding antiquarks, each in three varieties of "color," six leptons and their six corresponding antiparticles, one photon, three weak vector bosons, eight gluons and perhaps a graviton.

The standard model is based mainly on data from the large proton synchrotrons at Fermilab and at CERN, the European laboratory for particle physics, from the Stanford Linear Accelerator Center (SLAC), from the electron-positron colliders at Cornell, Hamburg and Stanford and, recently, from the proton-antiproton collider at CERN. Delicate low-energy experiments have also made important contributions to understanding. The quarks were originally introduced purely as theoretical entities after hundreds of hadrons had been discovered, in order to restore some underlying organization to the proliferating "elementary" particles. Quarks acquired a certain shadowy reality from the results of a large variety of experiments; it was not until 1974, however, that belief in their existence was firmly established by the simultaneous discovery at SLAC and at Brookhaven of the *J*/psi particle, which had been predicted on the basis of the quark hypothesis.

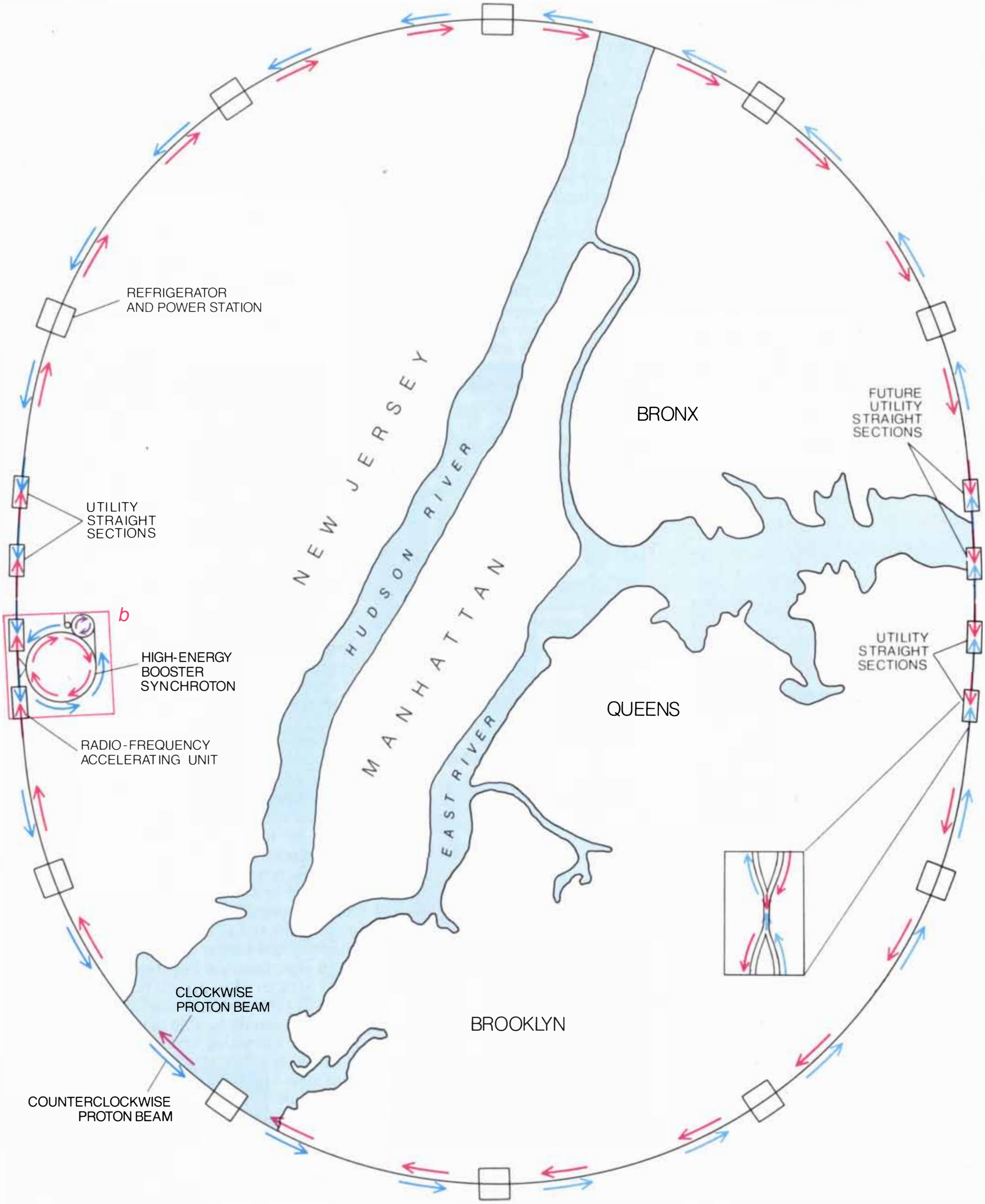
A central component of the standard model is the electroweak theory. In the present versions of that theory the six quarks and the six leptons are grouped into three generations; a pair of quarks and a pair of leptons are assigned to each generation. The electromagnetic and the weak interactions are described as different aspects of one underlying interaction called electroweak. The electroweak theory makes precise predictions about a great variety of phenomena, and it has been confirmed in detail by many experiments. Its most spectacular confirmation came at CERN in 1983 with the detection of the three weak vector bosons, the W^+, W^- and Z^0 particles.

The electroweak theory maintains a tradition that has characterized scientific thinking since its origins in ancient Greece: the unification of diverse phenomena under a single set of concepts. Indeed, to many physicists it is the paradigm of the kind of theory that may one day succeed in giving a unified account of all four fundamental interactions in nature. According to the electroweak theory, the unification of the weak and the electromagnetic interactions is manifest only at extremely high energies. At such energies the interactions are equivalent because the masses of the gauge bosons that mediate the two interactions are effectively zero. The full symmetry of the two interactions can come into play without inhibition.

The hypothesis of such a symmetry at high energies contrasts sharply with the properties of the two interactions in the ordinary laboratory environment. There the range of the weak interaction is roughly 1,000 times small-



SCHEMATIC MAPS OF THE SSC portray the machine at three scales: the central office and laboratory complex (*a*), the cascading series of three synchrotrons that preaccelerate the protons (*b*) and the main ring of the collider, superposed on an outline of the New York metropolitan area drawn to the same scale (*c*). Hydrogen atoms are ionized and the protons that make up their nuclei are accelerated (*purple*) by a linear accelerator to an energy of .6 GeV. The protons then enter the synchrotrons and are accelerated to an ener-


REFRIGERATOR AND POWER STATION
UTILITY STRAIGHT SECTIONS
b
HIGH-ENERGY BOOSTER SYNCHROTON
RADIO-FREQUENCY ACCELERATING UNIT
CLOCKWISE PROTON BEAM
COUNTERCLOCKWISE PROTON BEAM
NEW JERSEY
HUDSON RIVER
MANHATTAN
EAST RIVER
BRONX
QUEENS
BROOKLYN
FUTURE UTILITY STRAIGHT SECTIONS
UTILITY STRAIGHT SECTIONS

gy of 1 TeV in three stages; the beam is sent alternately clockwise and counterclockwise (*red, blue*) around the high-energy booster synchrotron for injection into the main collider rings. In a main ring each beam is further accelerated to 20 TeV. The protons are then made to collide and the collision by-products are monitored (*see inset to* c). Superconducting magnets are cooled by liquid helium surrounded by liquid nitrogen, both of which are pumped outward along the main ring from 10 refrigerator and power stations.

er than the diameter of the atomic nucleus, whereas the range of the electromagnetic interaction is infinite. According to the electroweak theory, this difference is a consequence of the fact that the weak gauge bosons are very heavy particles, whereas the mass of the electromagnetic gauge boson (the photon) is zero. The symmetry of the two interactions is said to break.

The Origin of Mass

Why is the symmetry of the weak and the electromagnetic interactions so badly broken at ordinary energies? The question becomes even more compelling because in the theoretical formulation of electroweak symmetry both the photon and the weak vector bosons initially have zero mass. Thus the observed heavy masses of the weak vector bosons arise out of the electroweak symmetry breaking, and the above question becomes in effect: What is the origin of mass?

The origin of mass is a problem central to determining the fundamental capabilities of the SSC. The analysis of electroweak symmetry breaking leads to a variety of potential theoretical scenarios, but they all share a common property: the evidence for or against any of them must become manifest at the collision energies now proposed for the SSC. There are, to be sure, many other questions arising from issues both within the standard model and beyond its scope that will be addressed by the SSC, and there are almost surely entirely unknown physical processes whose discovery the SSC will make possible. Nevertheless, most of the guidance we are able to get from theory as to the proper scale of the SSC comes primarily from considering the problem of the origin of mass.

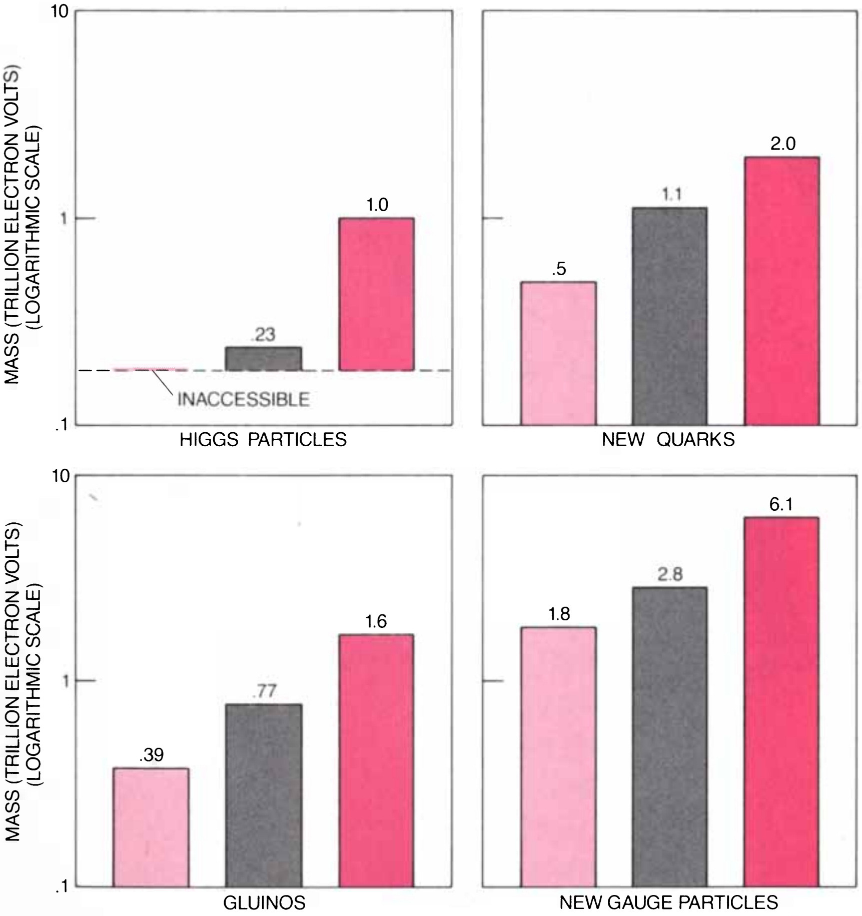
In the simplest version of electroweak dynamics the spontaneous symmetry breaking arises out of an electrically neutral field called the Higgs field, named after Peter W. Higgs of the University of Edinburgh. The Higgs field, if it exists, must assume a uniform, nonzero background value even in the vacuum. The idea that the vacuum "contains" anything, even a uniform, nonzero field, runs counter to the popular notion of the vacuum as empty space. In quantum mechanics, however, the air of paradox about such a result has been dispelled for some time. The quantum-mechanical vacuum constantly fluctuates with activity, whether or not the Higgs field is real.

The interaction of a particle with the Higgs field contributes to the energy of the particle with respect to the vacuum. That energy is equivalent to a mass. In the simplest model of the Higgs field the masses of the quarks, the leptons and the weak vector bosons are all explained as a result of the interaction with a single Higgs field. There is always a particle associated with a quantum-mechanical field, and so in the simplest form of the Higgs mechanism for symmetry breaking there is one Higgs particle associated with one Higgs field. If the Higgs particle exists, it should be possible to detect it, but searches so far have turned up nothing.

The Higgs Particle

One problem is that the mass of the Higgs particle is virtually unconstrained by theory. It could plausibly be as light as a few billion electron volts (GeV) or as heavy as 1 TeV. If the mass is less than 50 GeV, it should be found at the electron-positron colliders expected to be built and running by the end of this decade, such as the Stanford Linear Collider (SLC) or the Large Electron-Positron Collider (LEP) at CERN. If its mass is between 50 and 200 GeV, the Tevatron collider at Fermilab should be able to produce it, although extracting firm evidence for it amidst a welter of other particles in that mass range may be experimentally difficult.

For Higgs masses much greater than 200 GeV it becomes less appropriate to describe the Higgs as an elementary particle. Quantum mechanics teaches that the shorter the lifetime of a particle is, the less certainty there is about its energy, or equivalently its mass. If the Higgs mass is greater than 200 GeV, it decays into two W or two Z particles, and its lifetime must be so short that its mass is effectively



DISCOVERY LIMIT is the largest mass a new, hypothetical particle can have if it is to be detected by a colliding-beam accelerator of a given beam energy and luminosity. The graphs show how the discovery limits depend on accelerator beam energy and luminosity for four kinds of hypothetical particles. The graphs also show the lower limit for the mass of the Higgs boson: .18 TeV. A discovery is defined as the creation of 10 or more uniquely identified events in one year of data taking. The light-color graph for each particle gives the discovery limit for a proton-antiproton collider whose beam energy is 6 TeV and whose luminosity is 3 × 10^{31} in standard units. Such a collider could be built with existing technology in the 27-kilometer tunnel of the Large Electron-Positron ring at CERN, the European laboratory for particle physics. The light gray graphs give the discovery limits for a 20-TeV proton-antiproton collider whose luminosity is also 3 × 10^{31}. The dark-color graphs show the discovery limits for the SSC: a 20-TeV proton-proton collider whose luminosity is 10^{33}.

"smeared out" over a broad range. One is left to ponder what sense there is in regarding as a particle an entity that lacks definite mass.

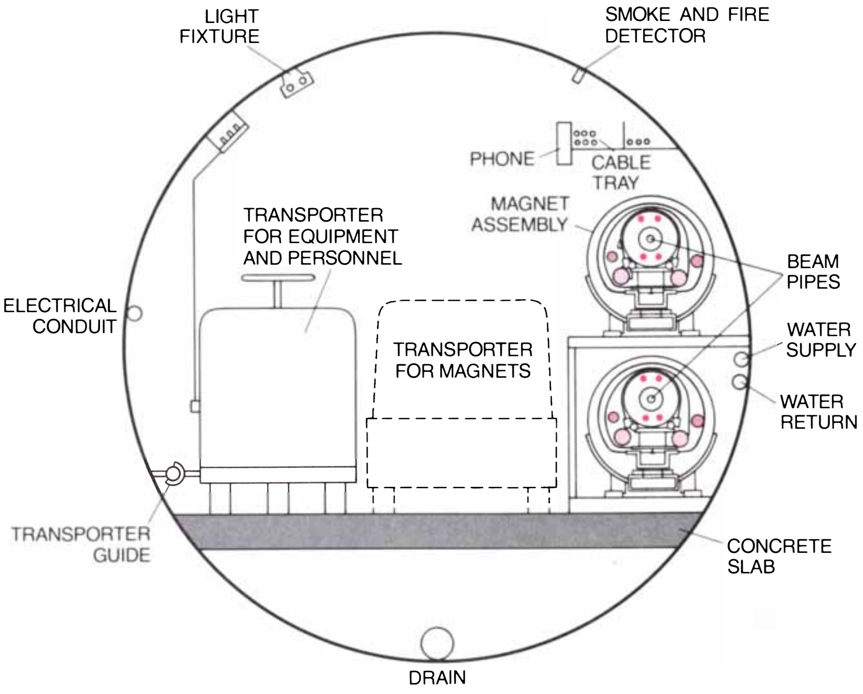
If the mass of the Higgs is as great as 1 TeV, the electroweak theory predicts that entirely new phenomena must emerge at energies of 1 TeV or more. In such circumstances the electroweak interaction becomes strong. Since electroweak dynamics governs the interactions of leptons and quarks, such particles may combine at energies of 1 TeV or more into composite particles with startling new features. In any case, if one is to confirm or disprove the Higgs mechanism throughout the potential mass range of a recognizable Higgs particle, the SSC will be needed.

Many physicists think the simplest form of the Higgs mechanism for symmetry breaking is only a low-energy approximation to reality. One reason stems from the fact that the Higgs particle, if it exists, cannot have a spin. If the Higgs particle had a spin, the Higgs field would have spin as well, and the mass of an ordinary particle would depend on its orientation in the vacuum. No such rotational dependence has ever been observed.

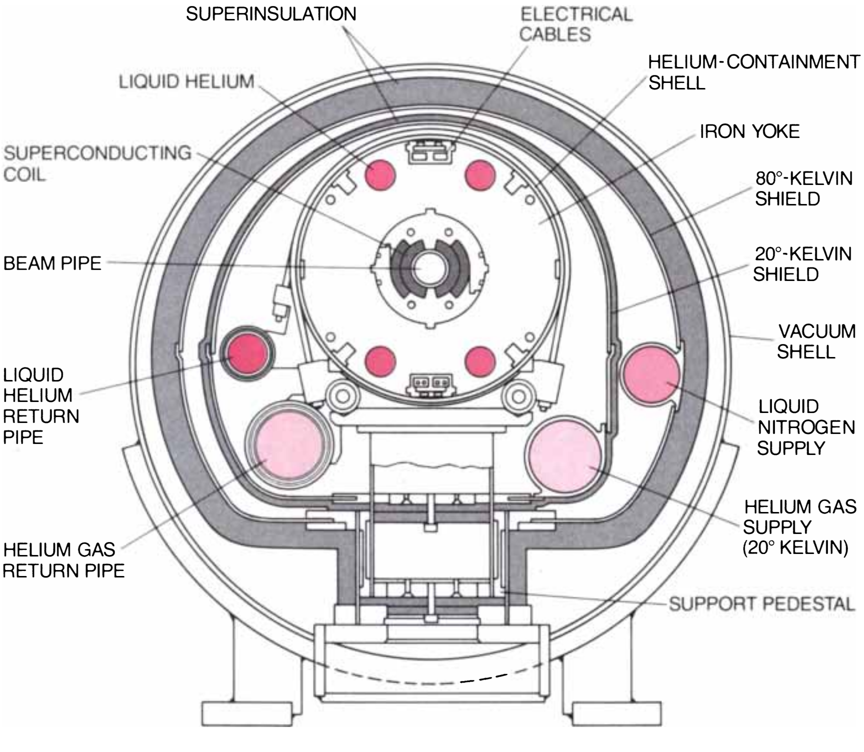
In quantum mechanics the spin of a particle can take on only discrete values, and particles that have integral spins (0, 1, 2 and so on) are sharply distinguished from particles that have odd-half integral spins (1/2, 3/2 and so on). Particles that have integral spin are called bosons, and so the spin-0 Higgs particle is a boson, just like the observed spin-1 gauge bosons such as the photon that mediate fundamental interactions. Particles having odd-half integral spin are called fermions, and they include all the quarks and leptons.

To calculate the mass of the Higgs boson one must make certain assumptions about physical processes at high energies. If the Higgs boson is an elementary particle, its calculated mass varies widely even for small changes in such assumptions. Such mathematical sensitivity has no natural physical interpretation; furthermore, it is not a characteristic of the expressions for the masses of spin-1/2 particles. Hence some theorists have proposed that one can avoid the mathematical difficulty and retain the necessary spin-0 Higgs boson if the boson is a composite particle instead of an elementary one, made up of two spin-1/2 fermions. Such composites are common in other domains of physics: the spin-0 pion, for example, is a composite particle made up of two spin-1/2 quarks. The spins cancel to give a composite without spin because the two quarks spin in opposite directions.

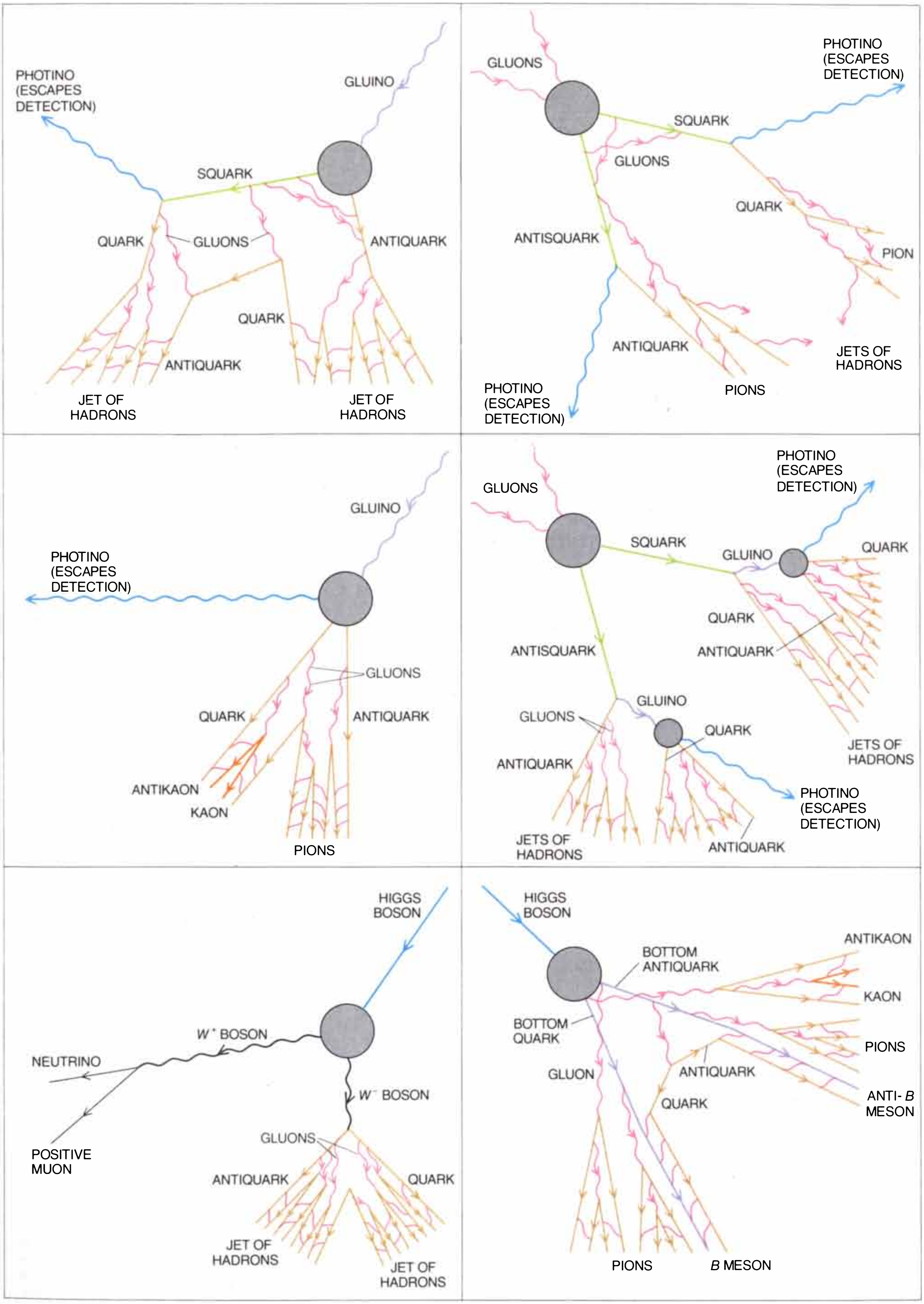
A composite Higgs boson would re-



MAIN TUNNEL proposed for the SSC is shown in cross section. The two beam pipes for the counterrotating beams of protons are at the right. The tunnel is about 10 feet across.



DETAIL OF MAGNET ASSEMBLY to be mounted in the tunnel of the SSC is shown in schematic cross section. One of the proton beams passes through an evacuated beam pipe in the central, upper part of the assembly. The pipeline is surrounded by coils of superconducting wire; current passing without resistance through the wire creates the enormous magnetic field needed for bending the proton beam. The rest of the system enclosing the beam pipe serves to keep the magnet at the low temperatures necessary to maintain the superconductivity of the coils. The first layer of piping surrounding the magnet carries liquid helium refrigerant held at 4.35 degrees Kelvin; this layer is surrounded by piping that carries liquid nitrogen at 80 degrees K. Layers of insulating material surround the piping.


PHOTINO (ESCAPES DETECTION)
GLUINO
SQUARK
QUARK
GLUONS
ANTIQUARK
QUARK
ANTIQUARK
JET OF HADRONS
JET OF HADRONS
GLUONS
PHOTINO (ESCAPES DETECTION)
SQUARK
GLUONS
QUARK
ANTISQUARK
PION
ANTIQUARK
JETS OF HADRONS
PIONS
PHOTINO (ESCAPES DETECTION)
GLUINO
PHOTINO (ESCAPES DETECTION)
GLUONS
QUARK
ANTIQUARK
ANTIKAON
KAON
PIONS
GLUONS
PHOTINO (ESCAPES DETECTION)
SQUARK
GLUINO
QUARK
QUARK
ANTIQUARK
ANTISQUARK
GLUINO
GLUONS
QUARK
ANTIQUARK
JETS OF HADRONS
PHOTINO (ESCAPES DETECTION)
JETS OF HADRONS
ANTIQUARK
HIGGS BOSON
W^+ BOSON
NEUTRINO
W^- BOSON
POSITIVE MUON
GLUONS
ANTIQUARK
QUARK
JET OF HADRONS
JET OF HADRONS
HIGGS BOSON
BOTTOM ANTIQUARK
ANTIKAON
KAON
BOTTOM QUARK
PIONS
GLUON
ANTIQUARK
QUARK
ANTI-*B* MESON
PIONS
B MESON

quire the existence of an entire new family of heavy, spin-1/2 particles called techniquarks. Such particles would be subject to a new strong interaction called the technicolor interaction, understood by analogy with the strong, color interaction that binds quarks into hadrons. Not only would techniquarks bind to form the Higgs boson but also they would bind to form a plethora of other composite techniparticles such as technipions, technivector mesons and so on. Such new particles would be quite heavy, but at least some of them must have masses of roughly between 50 and 500 GeV. At accelerators such as the Tevatron the number of such particles generated would be small, and they would be hard to detect against the background even if their masses are near the low end of the expected range. To test the theory one must employ a collider with a beam energy higher than several TeV, such as the SSC.

Supersymmetry

Another theoretical program with many attractive features is called supersymmetry, and it could provide an alternative to the simple Higgs mechanism for explaining the origin of mass. In a supersymmetric world every particle, including the Higgs boson, has a partner identical in every way except in its spin. To every ordinary fermion there corresponds a supersymmetric, spinless boson; for example, the spin-1/2 electron and quark have the spin-0 partners selectron and squark respectively. To every ordinary boson there corresponds a supersymmetric, spin-1/2 fermion; for example, the supersymmetric partner of the spin-1 photon is the spin-1/2 photino, the partner of the spin-1 gluon is the spin-1/2 gluino and the partner of the spin-0 Higgs boson is the spin-1/2 Higgsino.

If the supersymmetric particles existed in nature as exact copies of their counterparts except for spin, most of the supersymmetric particles would by now have been seen in abundance. Many searches have been made, however, and no evidence for the supersymmetric partners has been found. One might therefore suppose the appeal of supersymmetry would be on the wane, but that appeal persists for a number of reasons. One is that the existence of the supersymmetric partners would solve the problem of mathematical sensitivities in the theoretical expression for the mass of the Higgs boson. A second reason not to abandon supersymmetry is that it may be a broken symmetry in our world, just as electroweak symmetry is. A broken supersymmetry might give rise to supersymmetric particles that are substantially heavier than their ordinary partners.

No one clearly understands how sensitive mass is to symmetry breaking. For example, it is already known experimentally that the mass of the selectron, if it exists, must be at least 40,000 times as great as the mass of the electron. Does this ratio imply that supersymmetry must be "badly" broken? No one knows. What is known is that if supersymmetry turns out to be correct, it too, like the technicolor theory, introduces a new world of particles. Most of them must be quite massive; if they were not, they would already have been detected. Powerful new accelerators will undoubtedly be needed to find them.

There is a third and more general reason not to abandon supersymmetry, or for that matter any other theory such as technicolor that has a chance of explaining the mysteries of electroweak symmetry breaking and the origin of mass. No matter whether the Higgs boson is composite or elementary, whether or not it is embedded in a supersymmetric family of particles or indeed whether or not it exists at all, a general quantum-mechanical principle guarantees that new physical phenomena, deeply related to the origin of mass, should begin to emerge at energies of about 1 TeV. These phenomena must arise because if the existing standard model is extrapolated without any corrections into that energy domain, the probabilities calculated by the theory for certain interactions become greater than 1. Because no real probability can be greater than 1, the theory as it stands cannot be complete.

Since the correct theoretical extension of the standard model to very high energies is not known, the exact nature of the new physical phenomena cannot yet be described. If the Higgs boson is quite massive, one possibility already mentioned is that the electroweak interaction becomes a strong one. On the other hand, if the Higgs boson turns out to be light, its small mass could well be explained by supersymmetry. In that case the energy domain of roughly a few TeV would abound with the supersymmetric partners of known particles. The ability to probe that energy domain is therefore an extremely important goal for the basic understanding of matter.

The 20-TeV Requirement

Theories such as technicolor and supersymmetry yield specific predictions for the discovery limit of a collider that has a given energy and luminosity, or in other words the largest mass a hypothetical particle can have if it is to be created and detected at the collider. One might think a machine that brings about collisions of protons with a total energy of 40 TeV would make available roughly that amount of energy for the creation of new particles. Unfortunately only a fraction of the collision energy is actually released. All hadrons such as protons and antiprotons are composite systems, each one rather like a sack of marbles. The total energy is divided among the quarks, antiquarks and gluons that make up each hadron, and a collision can release only roughly the amount of energy carried by any two colliding constituents. For example, the Tevatron, with a total energy of 1.8 TeV, can thoroughly explore a range of masses up to only about .3 TeV.

To give some sense of how the largest detectable mass varies with collider design, consider the possible existence of the gluino and the squark. For a proton-antiproton collider whose beams are each accelerated to an energy of 6 TeV and whose luminosity is 3×10^{31} in standard units, the heaviest detectable gluino or squark would have a mass of about .4 TeV. With existing technology such a collider could be constructed in the LEP tunnel at CERN. At the SSC, where the luminosity would be increased over such a LEP collider by a factor of 30 and the energy increased by a factor of three, the heaviest detectable gluino or squark would have a mass of about 1.5 TeV. Similarly, the SSC could detect Higgs bosons as heavy as 1 TeV, new quarks as heavy as 2 TeV and new gauge particles as heavy as 6 TeV [*see illustration on page 70*].

The numerous theoretical estimates and machine assumptions can be challenged in detail, but the message is

SIX HYPOTHETICAL INTERACTIONS at the SSC could lead to the discovery of some of the new elementary particles now postulated by theory. The gluino and the squark are the so-called supersymmetric partners of the gluon and the quark. If both particles exist and the mass of the gluino is greater than the mass of the squark, the particles could decay as is shown in the top two diagrams. If the mass of the squark is the greater, the particles could decay as is shown in the middle two diagrams. In both cases the gluino decay is shown at the left. Two kinds of decay of the Higgs boson are shown in the diagrams at the bottom. The process at the bottom right would predominate if the mass of the Higgs boson were about 50 GeV; the one at the bottom left would be typical if the mass were 200 GeV or more.

clear. To probe effectively into the range of particle masses on the order of 1 TeV, a collider with the design characteristics of the proposed SSC is prudent. Less energy or less luminosity begins to compromise the discovery potential. Moreover, if the maximum available energy lies just below the threshold for the onset of some radically new physics, that physics will go undiscovered no matter what the luminosity is. The need for the highest feasible beam energies is paramount.

Proton Path

In order to understand how the SSC will achieve its design characteristics, it will be helpful to trace the path of the protons through the collider from their source to the sites of their collisions [*see illustrations on pages 68 and 69*]. The protons begin their journey as the nuclei of ionized hydrogen atoms in a gas. They are extracted from the gas by suitably arranged electrodes and emerge with a kinetic energy of a few thousand electron volts.

From there they enter a linear accelerator, a sequence of electrodes that accelerates the protons in a series of small pushes. In effect the protons are carried along by a precisely timed wave of potential difference that moves across the electrodes. The acceleration elevates the beam energy to 600 million electron volts. The beam then enters the first in a cascade of four synchrotron rings.

In synchrotron acceleration a uniform magnetic field forces the protons to follow a predetermined path. In a special section of the synchrotron ring the beam passes through a linear accelerator, and its energy is thereby increased with each circuit of the ring. As the energy increases, the strength of the magnetic field is also increased in order to keep the protons in their closed orbit. The frequency of the traveling wave that accelerates the protons within the linear-accelerator section must be synchronous with the frequency at which the protons circle the ring; the frequency is in the radio region of the electromagnetic spectrum. Modern acceleration systems are so efficient that the accelerator section will occupy a length of only 75 feet along the entire 52-mile circumference of the SSC.

The cost of the magnets that steer the beam is proportional to the diameter of the region in the beam pipe in which a uniform magnetic field is needed. The necessary field diameter depends on the diameter of the beam, and the beam diameter depends on, among other things, the ratio of the momentum of the particles in a direction transverse to the beam to their momentum along the beam path. The ratio decreases continuously because the acceleration system increases the momentum of the particles only along the beam path. In a large synchrotron the decrease in the beam diameter is exploited to minimize the diameter of the uniform magnetic field needed at each stage in the acceleration and thus minimize the overall cost. In our current design a cascade of synchrotrons sends the protons through a series of progressively narrower beam pipes. The first synchrotron will accelerate the protons from an energy of .6 GeV to about 8 GeV, the second will boost them from 8 to about 100 GeV and the third will boost them from 100 GeV to 1 TeV. The main, large synchrotron ring of the SSC will accelerate them to their final energy of 20 TeV.

At full energy the diameter of the beam will be a fifth of a millimeter. To enhance the luminosity of the collider the beams will then be focused at the collision points by powerful magnetic lenses into even tighter bunches. After focusing the bunches will be slim cylinders about 10 micrometers in diameter and 15 centimeters long. Each bunch will carry about 10 billion protons, and so the density of the bunch will be roughly one ten-thousandth the density of the molecules in the air at ordinary temperature and pressure. Because the probability of a collision between the protons in two counterrotating bunches is small, the bunches can interpenetrate repeatedly at the interaction points for many hours without need of replenishment.

Superconducting Magnets

The need to bend and focus the proton beam makes the magnet system a key element of any synchrotron, and in the past two years a considerable national effort has been devoted to studies of different kinds of magnets that might be adopted at the SSC. In principle it would be possible to build the SSC with copper-conductor electromagnets such as the ones used in the synchrotrons of the 1950's and 1960's. Such a ring, built to accelerate protons to an energy of 20 TeV, would consume at least four billion watts of power and lead to impractically high operating costs. Moreover, the magnetic properties of iron and the capacity of copper to carry electric current limit ordinary electromagnetic fields to a strength of about 2 teslas, or 20,000 gauss, which is roughly equivalent to the fields generated in the electric motors of home appliances.

Superconducting magnets can minimize both problems. They can sharply reduce the total power consumption of the acceleration system, and they can create magnetic fields several times stronger than conventional magnets can. A stronger magnetic field makes it possible to confine the protons to a tighter orbit for a given energy, and so it reduces the required size of the synchrotron ring.

In the conventional electromagnet wound with copper wire power is dissipated in forcing electric current through the wire. When certain metals and metal alloys and compounds are cooled to below a certain critical temperature, however, they become superconducting: they carry electric current without resistance. In a magnet wound with superconducting wire the only power required is that needed to maintain the wire below the critical superconducting temperature. The refrigerators needed for the SSC would consume about 30 million watts of power, which is somewhat less than the total required by the largest accelerator complexes today.

In 1983 the first superconducting synchrotron, the Tevatron, went into service. Its successful operation since then has provided data that have been invaluable for planning the SSC. Last September we and our colleagues in the SSC Central Design Group reached a milestone by making our final choice for the design of the superconducting, bending magnets. The bending magnets will have a field strength of about 6.6 teslas, and that strength will give a circumference of about 52 miles for the main synchrotron ring.

Both the superconducting coil of the magnet and a surrounding yoke of iron will be held at a temperature of 4.35 degrees Kelvin by a pressurized flow of liquid helium. Ten refrigerators will be spaced uniformly around the ring, and each one will supply a coolant stream along the ring out to about four kilometers in both directions. An intermediate heat shield of liquid nitrogen, maintained at 80 degrees K., will sharply reduce the heat energy that is incident on the liquid helium. To stabilize the entire system against temperature changes and mechanical vibrations the ring will be housed in a tunnel. The tunnel will then be covered with at least six meters of earth in order to safely absorb any ionizing radiation that might be generated if the beam were to strike the walls of the beam pipe.

Detection Apparatus

Because the SSC will explore a previously unknown energy regime, no one can predict with certainty the properties of the most interesting events.

Hence it is essential to maximize the variety and flexibility of the particle detectors planned for the machine. Experiments with the initial complement of detectors will guide the ultimate experimental program.

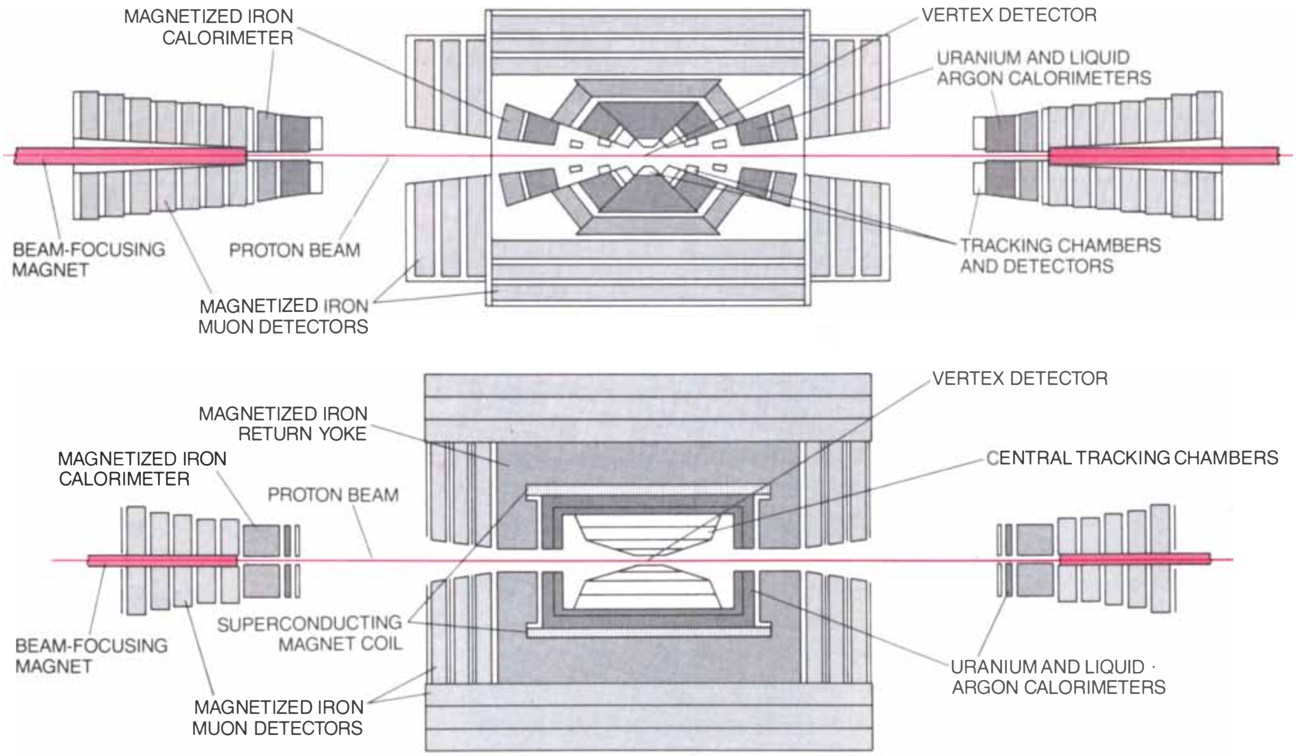
The collisions between the counterrotating beams of protons will be generated by forcing them to intersect head on, or at a small angle, at one of several points along the ring. At such interaction points there will be elaborate detectors housed in experimental halls. The detectors will record the passage of particles emerging from the collisions and sift through them electronically in order to identify the ones of potential interest.

We expect to generate up to 100 million collisions per second in each interaction region of the SSC, and so the electronic sifting is not at all trivial: one must find a needle or at most a few needles in a haystack of data. Furthermore, once such a potentially interesting collision event is identified, the event must be subjected to additional, more detailed tests. If it still passes the tests, enough information must be collected and recorded about the tracks emerging from the collision to allow a later reconstruction of what actually took place. The entire process must be done as quickly as possible to minimize the "dead" time in which the detector and its associated computer systems are not recording collision data.

Since the need to minimize the dead time requires that one record only a small fraction of all the collisions, one must design an electronic trigger for the interesting phenomena. Here one is guided both by past history and by theoretical ideas. For example, the W^+, W^- and Z^0 bosons were signaled by the emission of energetic electrons and muons and, in the case of the two W bosons, by the emission of neutrinos. The Higgs boson is expected to give rise to decay signatures similar to those of the weak bosons. Theory predicts it should decay into the most massive particles possible, such as heavy quarks or W bosons. The quarks or W bosons should in turn give rise to electrons, muons and neutrinos. Neither the Higgs boson nor the heavy quarks and W bosons can live long enough to be directly detected, and so one must infer their presence from their by-products.

Because electrons and muons are electrically charged, their presence can be directly detected; the detection of the neutrinos, however, is subtler. The presence of the neutrino can be inferred by measuring the total momentum, in directions transverse to the proton beams, of the particles emerging from a collision. Since the total momentum of the colliding protons in directions transverse to the proton beams is essentially zero, the total transverse momentum of the emerging particles must also be zero. Thus the total momentum carried by particles in one direction transverse to the beam must be balanced by the total momentum carried in the opposite direction. If it is not, one can assume the imbalance was caused by the passage of a neutrino, a particle that interacts so rarely with matter that it is almost never detected.

The first trigger at the SSC might require only the passage of a moderately energetic electron or muon. Higher levels of decision making might then integrate increasingly sophisticated information from several kinds of detector about the energy and flight path of the emerging particles. For some kinds of triggering it may be sufficient to


MAGNETIZED IRON CALORIMETER
VERTEX DETECTOR
URANIUM AND LIQUID ARGON CALORIMETERS
BEAM-FOCUSING MAGNET
PROTON BEAM
TRACKING CHAMBERS AND DETECTORS
MAGNETIZED IRON MUON DETECTORS
VERTEX DETECTOR
MAGNETIZED IRON RETURN YOKE
MAGNETIZED IRON CALORIMETER
CENTRAL TRACKING CHAMBERS
PROTON BEAM
SUPERCONDUCTING MAGNET COIL
BEAM-FOCUSING MAGNET
URANIUM AND LIQUID ARGON CALORIMETERS
MAGNETIZED IRON MUON DETECTORS

GENERAL-PURPOSE PARTICLE DETECTORS at the SSC would be designed to cover the entire space surrounding the collision site of the two proton beams. The upper detector has no magnetic field around the central tracking chambers; the lower one incorporates a superconducting magnet that generates a field along the direction of the beam. Both detectors have a central part and a forward part; the forward part is designed to detect the particles moving almost parallel to the beam. The inner part of each detector is designed to measure various properties of each charged track; such properties can later be used to reconstruct the path of a particle. The outer part of each detector is a calorimeter, which measures the total energy deposited in each of its segments. The outermost subsystem in each detector is made up of magnetized layers of iron interspersed with tracking chambers; this subsystem identifies the muons and measures their energies. All other known particles except the neutrinos are completely absorbed by the calorimeter.

build a detector that surrounds only part of the collision site. In most cases, however, general-purpose detectors are needed that surround the collision site as completely as possible. We can best illustrate the functions of a detector by describing the main features of the general-purpose detector.

General-Purpose Detectors

For a great many measurements a general-purpose detector with a magnetic field in its central region gives a decided advantage, but there is a price attached. The increased complexity of the tracking and the inert material needed for the magnetic coil make measurement hard, and the coil adds significantly to the cost. Hence an optimum design for the initial detectors at the SSC might include one detector with a magnetic field and another detector without one.

The components of such detectors are quite similar [*see illustration on preceding page*]. The innermost component will probably be a so-called vertex detector, designed to measure the tracks of particles as precisely as possible near the vertex, or site, of the proton collision. The particles then pass outward into a central tracking chamber that measures the direction and, if the magnetic field is present, the curvature of the tracks of particles that carry electric charge. Just outside the tracking chamber, or perhaps interspersed with it, there may be detectors that identify the electrons.

The next group of active detectors are the calorimeters, which measure the total energy of all the particles emitted in directions greater than some critical angle to the direction of the incident proton beams. One purpose of the calorimeters is to infer the presence of high-energy neutrinos that result from "hard," or head-on, collisions between the constituents of two protons. Remember that in principle the neutrinos can be detected by finding an apparent violation of the conservation of transverse momentum. For particles moving close to the speed of light, momentum is equivalent to energy. In practice, therefore, one can detect an imbalance in the transverse momentum of the collision by-products by detecting an imbalance in their transverse energy.

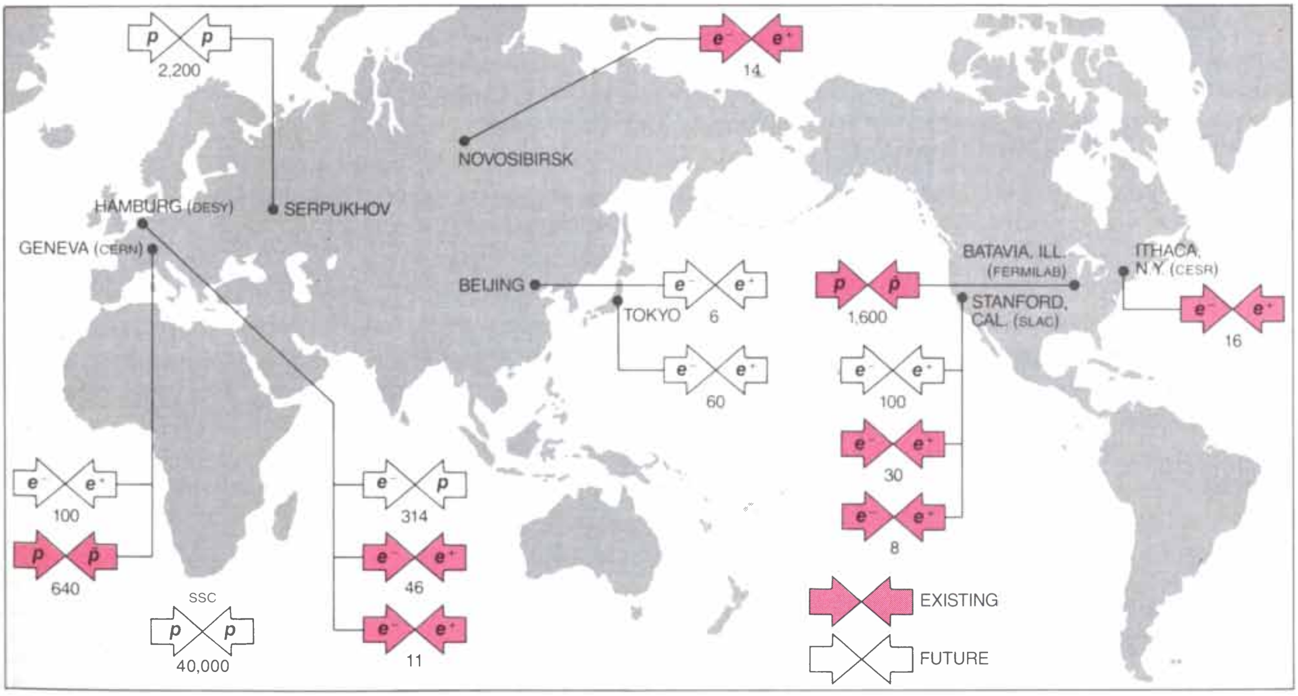
The measurement of transverse energy could be done by individually measuring the energy of each particle coming out of the collision. Such a measurement, however, would be prohibitively difficult. Each collision at the SSC could easily generate more than 100 particles, and many of the particles are expected to emerge in jets, or tight bunches, making them even harder to distinguish. Furthermore, a standard technique for determining the energy of a charged particle, which depends on measuring the curvature of its path in a magnetic field, cannot be applied to a neutral particle.

The calorimeters measure the total absorption of energy in some medium without distinguishing the separate contributions of individual particles. In spite of this limitation the direction of the energy deposits can be determined by segmenting the calorimeters. Thus the total transverse energy and the direction of jets can be measured quite well, and from the data one can determine whether or not a neutrino was generated by the collision.

The calorimeter in the general-purpose detector doubles as the first layer of the system for identifying muons, which are often important signals of interesting events. Outside it there will be several layers of magnetized iron, with tracking chambers interspersed among the layers. The aim in the design here is to allow for redundancy in the measurement of the momentum of the muons: relatively unimportant decays can readily simulate the passage of an energetic muon.

Frontiers of High-Energy Physics

We have stressed that the creation and detection of particles having masses above 1 TeV at the SSC will extend knowledge about elementary processes well beyond the compass of the standard model. In particular it will address the fundamental problem of the origin of mass and the problem of symmetry breaking in the electro-

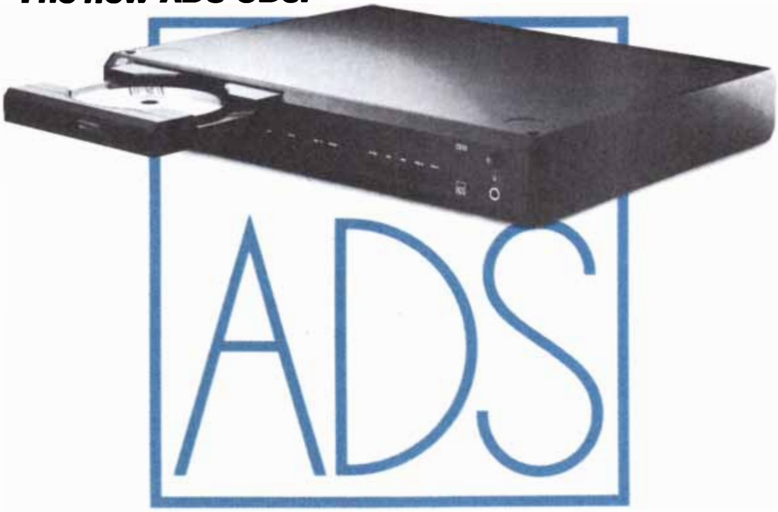


COLLIDING-BEAM ACCELERATORS now dominate the technology for high-energy experiments in physics. The map shows both existing and planned machines. The accelerators are labeled according to the particles they make collide: e^- for the electron, e^+ for the positron, p for the proton and $\bar{p}$ for the antiproton. The numbers give the maximum total collision energy in billions of electron volts.

weak theory. The SSC will also make fundamental contributions, however, to many other open questions. So far, for example, there appear to be three generations of quarks and leptons. Are there more? Why do the quarks and leptons have progressively greater masses in successive generations? Are the quarks always bound together to form hadrons or shall we ultimately see manifestations of free quarks? Are the quarks and leptons related, and if they are, how? Why do weak interactions show a handedness? Are quarks and leptons really elementary entities or are they built up from some more basic constituents? Does quantum mechanics continue to apply at smaller and smaller scales? Can gravity, as well as the color interaction, be treated in a consistent way by quantum mechanics and perhaps unified with the other known forces?

In the past decade there have been several attempts to extend the partial unification found in the electroweak theory to a grand unification of the electromagnetic, weak and color interactions. Even more recently a development called superstring theory has extended the theory of supersymmetry to a mathematical formalism that may one day bring about an even grander unification: the unified understanding of all four fundamental interactions, including gravity. Out of such grand unified theories has emerged a realization that particle physics has something to say about the earliest epochs in the history of the universe and that cosmology has something to say about particle physics.

Astronomers now believe the universe began cataclysmically in the big bang. In the almost unimaginably hot, primordial universe just after the big bang the full symmetry of nature's laws must have been manifest. Both the study of the very large and the study of the very small thus converge on a common point of view: in order to continue probing nature's underlying unity and simplicity one must build instruments that investigate domains of progressively higher energy. The discoveries in such domains cannot be fully anticipated, but experience teaches us it is often the unexpected discovery that triggers a deeper scientific understanding of the world. The SSC, ambitious yet feasible, would take us to domains of energy never before encountered, where the real discoveries can only be guessed at, and it would give us access to the events that took place almost immediately after the beginning of time. The opportunity and the challenge presented by the SSC will excite all who share our desire to understand the natural world.



Computer-simulated Plant Evolution

Computers are appropriate tools for testing the inherently statistical hypotheses of evolutionary biology. A desktop computer can re-create the major trends in plant evolution

by Karl J. Niklas

Evolutionary biologists encounter a fundamental difficulty when it comes to testing theories: the hypotheses on which the theories are based often do not make specific and easily falsifiable predictions. Instead they attempt to describe general statistical trends that should be discerned in large populations over long periods of time. In addition both the organisms in question and their environments may have vanished hundreds of millions of years ago—as in the case, for example, of the evolution of the earliest land plants. Therefore when the paleobotanist, the paleozoologist or the geneticist have drawn the evolutionary history of a lineage in great detail, it may still be impossible to give definitive answers to certain basic questions: Why do the observed patterns of evolution exist? How much of what is in the fossil record should be attributed to chance and how much to clearly defined biotic events and selective pressures?

There is, however, an effective tool for testing evolutionary hypotheses: the computer. It can handle large sets of data and do rapid and repeated calculations, and with it the investigator can model complex evolutionary processes. The techniques of computer modeling make it possible to examine many of the intuitive notions formed by biologists about the interactions among organisms and environments. It is this kind of work that has engaged my colleagues Vincent Kerchner and Thomas D. O'Rourke and me at Cornell University. We have been testing the mathematical consequences of various notions about plant evolution.

Computer simulations examine hypotheses by what has come to be known as the hypothetico-deductive method. The first step is to formulate a hypothesis. The various consequences of the hypothesis are then deduced and compared with observational evidence. If the consequences agree with the observations, the hypothesis is partially confirmed. (A hypothesis cannot be "proved" in this way, only made more probable.) If one of the deduced consequences of the hypothesis does not agree with observation, the hypothesis must be modified or rejected. The computer, which can produce a large "population of results" by repeating the same general type of computation many times, is an ideal tool for testing the inherently statistical hypotheses of evolutionary biology.

The first step is to formulate hypotheses: statements about what factors have had the greatest effect on plant evolution. The computer can then be used to model the performance of plants having various primitive characteristics and "score" the relative success of each simulated plant in solving the problems presented by the hypothesized selective pressures. Next, small or large changes ("mutations") that might make the primitive plants fitter, which is to say more able to cope with the hypothesized pressures, can be introduced into the simulated plants. The mutated progeny are themselves scored and are then allowed to mutate further. This process is repeated many times. The final step is to compare the pattern of evolution thus simulated with the patterns found in the fossil record. If the two are in good agreement, then the operation of the factors that were hypothesized to exert selective pressures is partially confirmed.

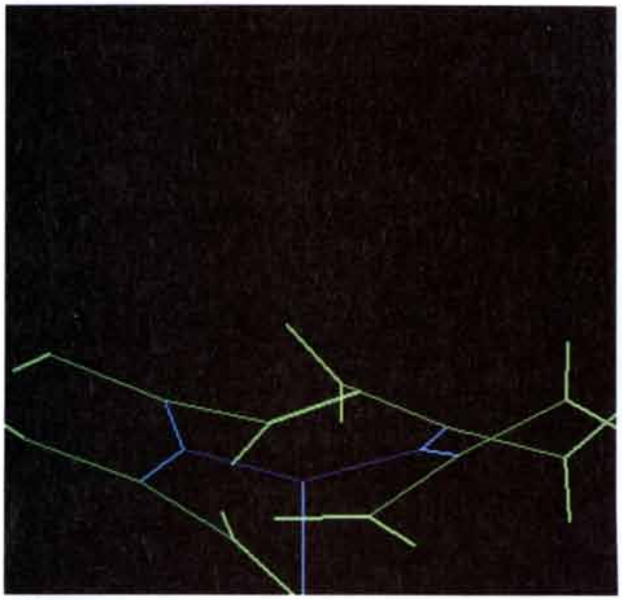
This modeling procedure is based on two major assumptions, which can be regarded as the two major tenets of evolutionary biology. The primary assumption is that the genetic character of individuals within a species, and hence of the species itself, changes with time. The second is that there is a degree of genetic continuity between an ancestor and its descendants. The changes within a species are therefore the result of selective pressures applied to many minor variations that arise between parent and offspring.

In order to model the evolution of plants some specific hypotheses must be added to these assumptions. One such hypothesis is that the majority of plants can be seen as structural solutions to constraints imposed by the biochemical process of photosynthesis. Plants with branching patterns that gather the most light can then be predicted to be the most successful. Consequently changes in the plant's shape or internal structure that increase its ability to gather light should confer competitive advantages.

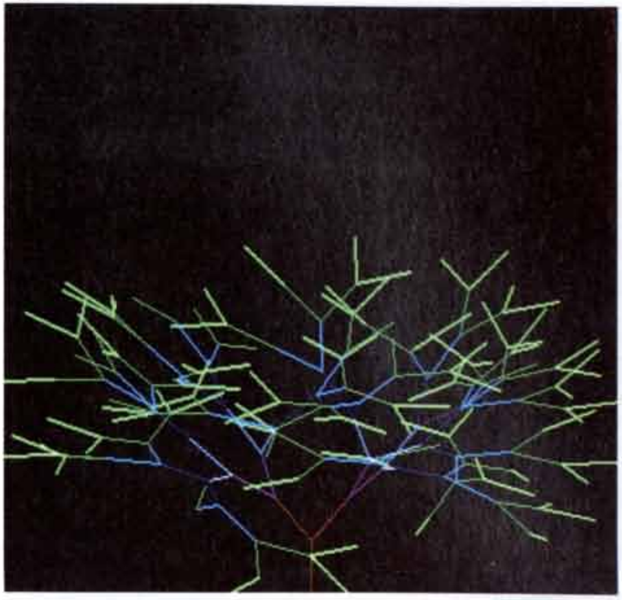
To be effective competitors for light

EVOLUTIONARY DEVELOPMENT of the early land plants is simulated on a computer. This simulated evolutionary sequence is based on the hypothesis that evolution was driven in part by a plant's need to minimize the mechanical stresses inherent in its branching pattern while exposing the maximum photosynthetic surface to the sun. In the simulated sequence, which exhibits some of the major trends seen in the actual evolution of plants, the most primitive plant (*1*) is low and sparsely branched. (Different colors indicate successive generations of branching; the light green branches are the ones most recently formed.) Later plants tend to have more branches (*2*) and to grow more vertically (*3*); vertical branching patterns present more photosynthetic surface and can grow above the shade cast by such obstructions as neighboring plants. The amount of shade the plant casts on itself is then reduced by the evolution of larger angles between branches, resulting in a plant with fuller growth (*4*). Later plants (*5*) have evolved a single central axis from which many lateral branches grow. The light-gathering ability of this configuration is increased by planation, or flattening, of the lateral branching systems (*6*). The similarity between the trends produced by such a simulation and the trends actually observed in the fossil record is a measure of the accuracy of the hypotheses on which the computerized simulation is based.

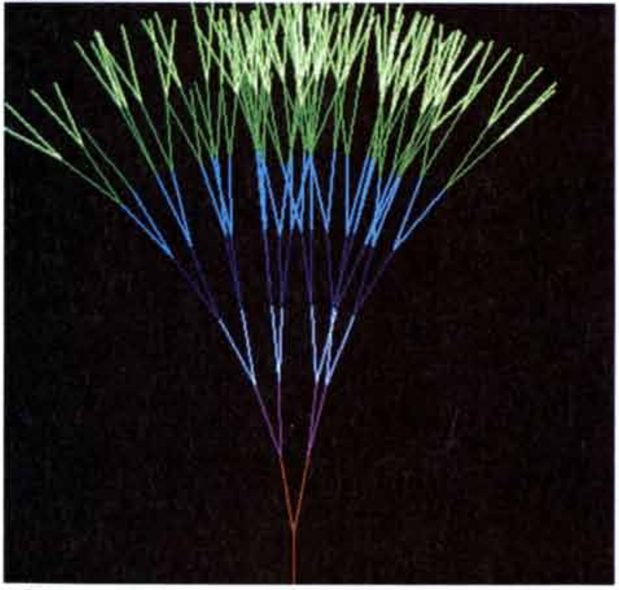
1



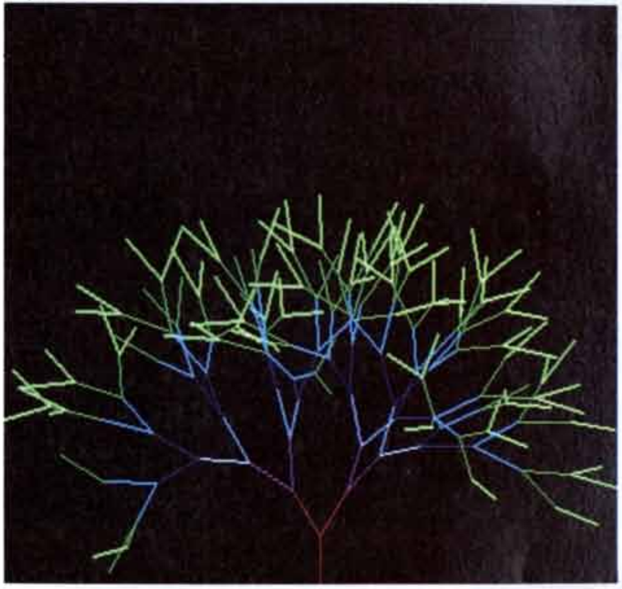
2



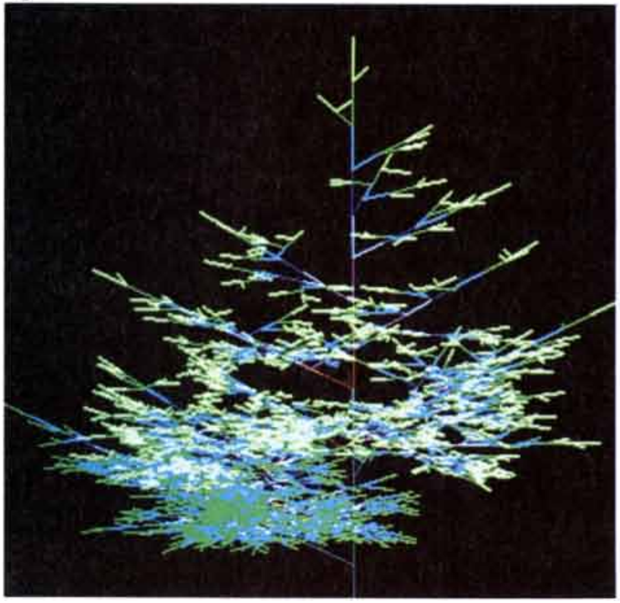
3



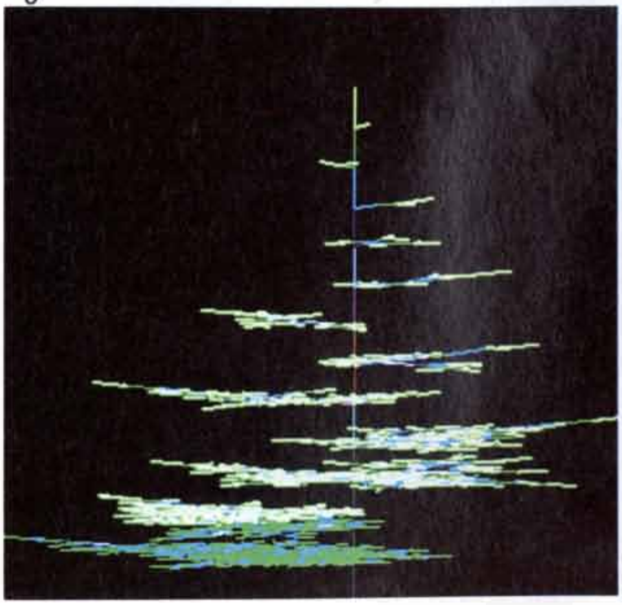
4

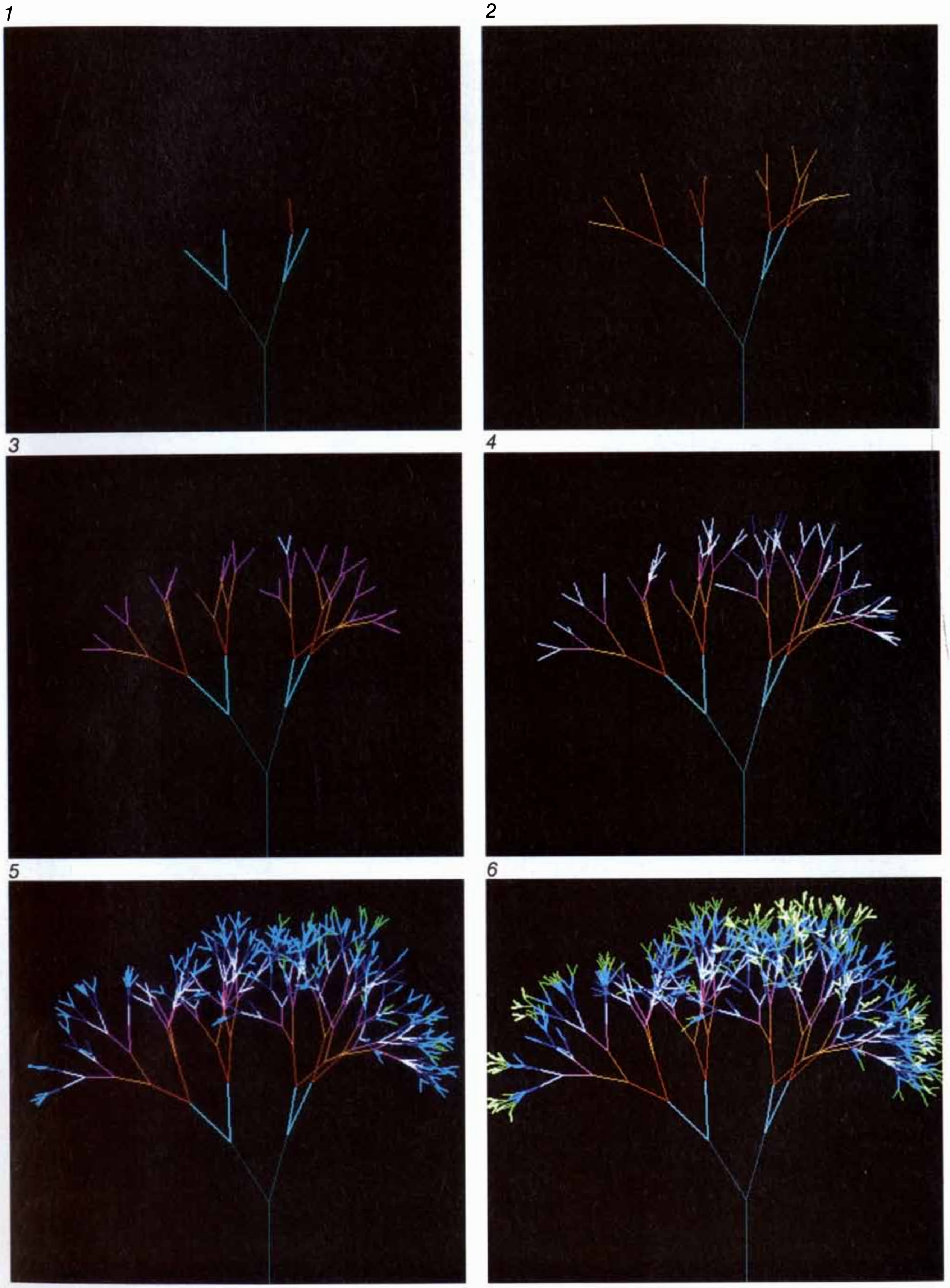


5



6





SIMULATED GROWTH of a plant consists of a series of iterated stages. In each iteration the computer decides, based on the plant's inherent "probability of branching," which of the plant's "axes" will branch. It then considers the plant's other inherent characteristics, such as the angles between branches, to determine the directions of the new axes, and it allows each to grow a short distance.

and space, plants must perform certain other tasks. In particular they must be able to stay erect: to sustain the mechanical stresses involved in vertical growth. A second hypothesis, then, might be that the evolution of plants was driven by the need to reconcile the ability to gather light with the ability to support vertical branching structures. A third hypothesis might be that the evolution of plants is driven by the extent to which they are successful at reproduction, placing a premium on branching patterns that allow for better dissemination of seeds or spores.

With this set of assumptions and hypotheses, many of the major trends seen in the early evolution of land plants can be simulated by a desktop computer. The simulations completed so far refer only to the first phase of the spread of vascular plants (plants containing internal tissues—xylem and phloem—that translocate fluids and also help to support the plants' structure vertically). The period simulated is only about 60 million years long, from about 410 million to about 350 million years ago. (In contrast, the diversification of the flowering plants, the most recent group of plants to evolve, took place over a period of about 100 million years.) Hypotheses related to terrestrial herbivores and pollinating insects have not been considered. Such hypotheses would be much more complex to model than those discussed here. With more powerful computers it may eventually be possible to model these features and others that contribute to evolution, such as the effects of climatic changes and catastrophic events.

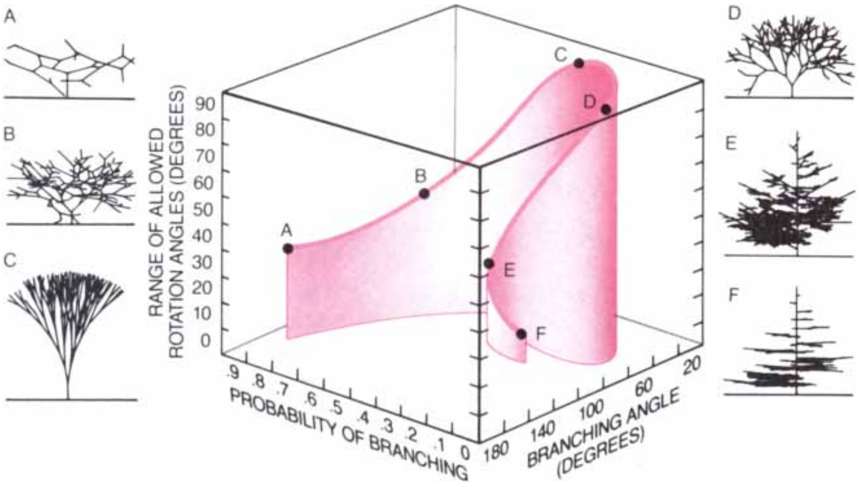
In order to simulate plant evolution one must develop mathematical techniques for quantifying the competitive advantages offered by various features. One such technique should determine how much a light a plant having a given structure could intercept. As far as is now known, the first vascular plants, which gave rise to the bulk of the earth's current flora, were leafless, with vertical photosynthetic axes. (Stems, which are axes bearing leaves, evolved later.) The axes tended to grow lengthwise, through the addition of new cells produced by a cluster of cells at an apex, or growing end; they generally added only a small number of new cells to their girth and so had a limited maximum diameter. Branching took place when a single axis bifurcated at its apex into two independently growing axes.

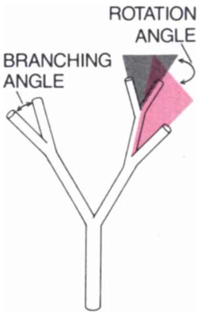
Because many of these plants lacked leaves, the relatively stiff axes were the chief photosynthetic organs. Hence the geometry of a plant's branching and the way the surfaces of its axes were projected toward the sun were the most important factors in determining the plant's ability to gather light. Unlike the leaves of today's plants, the axes of the first vascular plants could neither flutter appreciably in a breeze nor track the movement of the sun during the day. The light-gathering ability of the early plants can therefore be simulated by programs that determine the total amount of sunlight projected onto a static three-dimensional branching pattern as the sun's position in the sky changes.

Another parameter to quantify is the mechanical stability of vertical axes. Branching patterns that grow mainly in the vertical direction are more efficient at gathering light because they can reach beyond the shadows projected by such obstacles as rocks, hillocks and other plants, but a plant with a vertical posture must be able to sustain the concomitant mechanical loading. The principal load-bearing innovation of modern trees is the woody stem that increases in girth continuously as new cylindrical layers of cells are added internally. Without such secondary wood, the ability of the early vascular plants to sustain mechanical stresses depended solely on their primary growth: the relatively slender, virtually untapered axes produced when new cells were added only at the tips of preexisting axes.

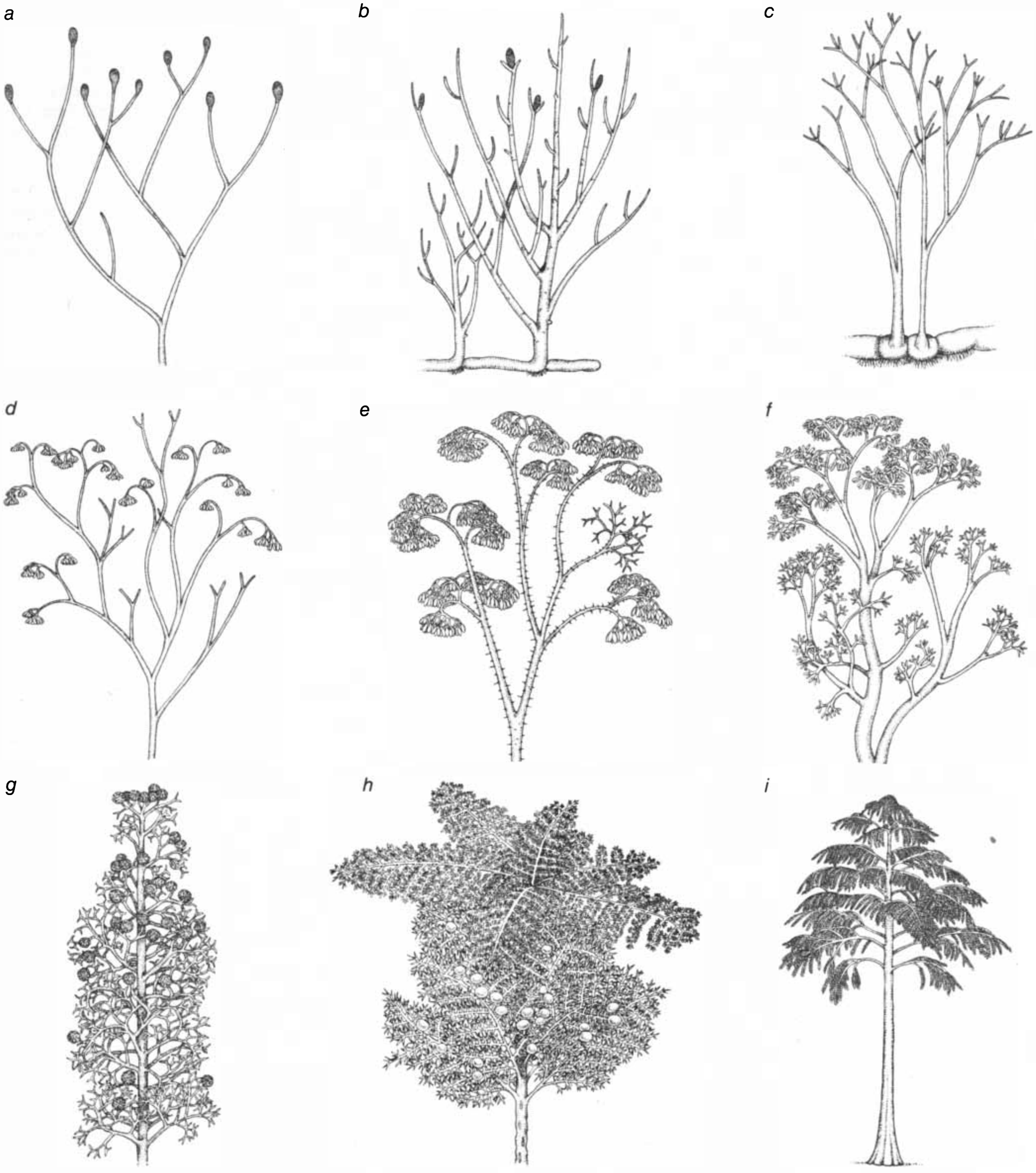
It is rather easy to calculate the amount of load inherent in a branching pattern, as well as that pattern's ability to sustain the load, if the axes' weights, sizes and orientations are known. The programs that generate branching patterns can also calculate the total strain and bending moment of a branching pattern, as well as the amount of sunlight intercepted. This makes it possible to quantify the factors involved in the tradeoff between, on the one hand, presenting large areas of photosynthetic tissue to the sun and, on the other hand, bearing the resulting mechanical stresses.

Another tradeoff to be considered has to do with shading. Larger and more extensively branched plants have a greater capacity to shade neighboring plants and so gain an advantage over them. This capacity has a negative aspect as well, however: it is likely to increase the plant's tendency to shade parts of itself, thereby reducing its own light-gathering efficiency.

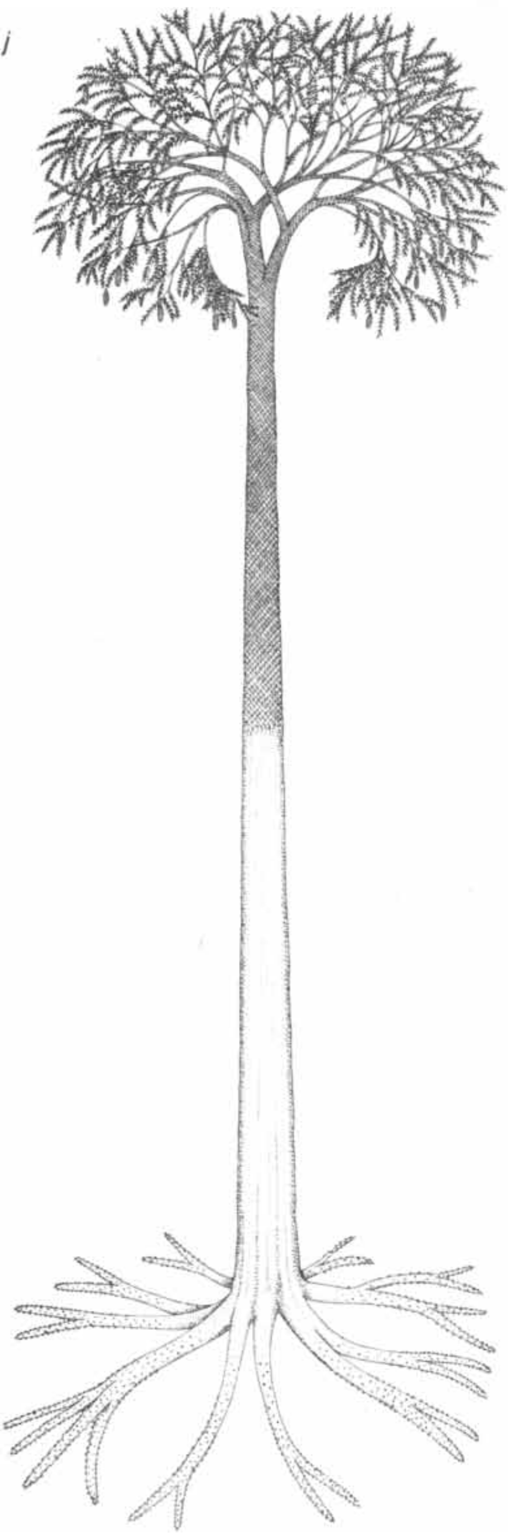




CUBICAL SPACE represents a "universe" of possible branching patterns. Each of the cube's three mutually perpendicular dimensions represents a particular morphological characteristic, and each point within the cube represents a "species" of plant with a unique combination of characteristics. The three characteristics represented are the probability of branching, the branching angle (the angle between adjacent branches) and the allowed range of rotation angles (the angle between the plane defined by a pair of new axes and the similar plane, formed during the previous generation of branching, defined by the axis from which the new axes have grown and that axis's sibling). The computer can simulate an "evolutionary trajectory" (*color*) within the cube. The computer starts with the plant that most nearly resembles a primitive species. It then examines each of that plant's near neighbors in the cube and determines which is the fittest according to hypotheses selected by the investigator. That neighboring plant is the next point on the trajectory, and the computer then scans all its near neighbors for a fitter plant. The procedure is repeated until the computer has located a plant that is fitter than any other plant in the universe of possible morphologies.



EVOLUTION OF ANCIENT PLANTS shows trends, such as an increase in the number of branchings, the emergence of a main axis and the planation of branching systems, that are also seen in simulated evolutionary trajectories. (Many of the plants depicted are not directly related to one another, and so they illustrate general evolutionary tendencies rather than the history of any particular lineage.) One of the earliest vascular land plants was *Steganotheca striata* (*a*), a sparsely branched plant that lived about 450 million years ago. *Rhynia gwynne-vaughanii* (*b*) and *Horneophyton lignieri* (*c*), both of which have more extensive branching structures than *Steganotheca* has, may have been among its descendants. All three plants are part of the rhyniophyte group. The next major group, which lived roughly 380 million years ago, was that of the trimerophytes, which included *Trimerophyton robustius* (*d*), *Psilophyton charientos* (*e*), *Psilophyton dawsonii* (*f*) and *Pertica quadrifaria* (*g*). (The exact taxonomic relations and sequence of occurrence of these trimerophytes are not known, and so they are depicted in a somewhat arbitrary sequence.) The trimerophytes seem to have a main vertical stem; in fact, this apparent main stem is the statistical result of a large number of "unequal branchings," in which one member of a pair of new axes grows in an orientation closer to that of the pair's parent axis. A fernlike descendant of the trimerophytes, *Rhacophyton ceratangium* (*h*), does indeed have a main vertical stem; it also has planated prefronds, leaflike appendages that mark



an advancement over planated branches. Another possible descendant of the trimerophytes is *Archaeopteris* (*i*). Plants of this genus had a central trunk bearing lateral branches, which in turn bore leaves. Still another lineage, that of the lycopods, seems to have evolved independently from the rhyniophytes and the trimerophytes, passing through many (but not all) stages characteristic of simulated evolutionary trajectories. Lycopods reached their apex in tree lycopods, such as the genus *Lepidodendron* (*j*).

The hypothesis that light-gathering ability was a primary driving force in the early evolution of vascular plants is certainly a simplification. There is a plethora of shade-tolerant plant species in today's flora, and there is reason to believe some early vascular land plants were shade-tolerant as well. Nevertheless, the fossil record does show long-term trends consistent with the hypothesis that competition for light was primary in the evolution of land plants. Among these trends are an increase in the stature of successively evolved plant groups, a transition to a growth pattern characterized by a single vertical axis bearing lateral axes and the appearance of planated, or flattened, lateral branching systems, which in some lineages mimicked the function of leaves.

These trends have been re-created with some accuracy by simulations in which branching patterns "grow" in ways defined by certain "genetic" characteristics. During each stage of simulated growth, a plant goes through several phases of alteration. First every axis grows a small distance. Then the computer selects on the basis of a predetermined genetic "probability of branching" how many (and which) axes will branch. On those that do branch, the directions of the new axes are determined by two "genetic" factors. One is the "branching angle" between the new axes. The other is the rotation angle: the angle between the plane defined by the two new axes and a plane formed by the previous generation of branching on the parent axis [*see illustration on page 81*].

In the simplest simulations of branching patterns the branching angle is the same at all branching points; like the probability of branching, it is one of the pattern's "genetic" characteristics. The rotation angle, on the other hand, may vary within a specified range; the range is a characteristic of the simulated plant. Plants that have a large allowed range of rotation angles tend to be full and bushy; in plants where the rotation angle cannot deviate very much from 0, groups of branches tend to fall on parallel planes. After the computer has been used to determine the location and direction of each new axis, all the new axes are allowed to grow a short distance and the process of branching is repeated. The plants are considered complete and their growth is terminated after 10 cycles of branching. A simulated plant in which every axis branches during each of the 10 cycles of growth contains 1,024 axial elements. Larger plants can be simulated, but this would require computers with larger memories. Ten cycles of branching are enough to mimic the size and shape of the majority of fossil plants under consideration.

The three characteristics—probability of branching, branching angle and rotation angle—define a universe of possible branching patterns that characterizes the morphology of the earliest land plants. The simplest way to imagine this universe is as a cube. Each of the cube's three mutually perpendicular dimensions represents one of the three basic characteristics, and so every point within the cube represents all the plants that have a particular set of attributes. (Because of the random nature of branching, plants having exactly the same mathematical characteristics may be slightly different morphologically.)

For example, suppose the vertical dimension of the cube represents the branching angle. Then any point near the top of the cube would represent a "species" whose branches diverge from one another at a large angle, whereas a point near the bottom of the cube would represent a species that has a small branching angle. One corner of the cube represents a "species" of plant with a branching probability of nearly 0, very small angles between branches and rotation angles that may not deviate much from 0. Points farther away from this corner in any of the three primary directions represent respectively plants that have more branches, those with greater branching angles or those whose branches grow in a greater variety of directions. The combinations of the three variables represented by the sites within the cube include virtually all the branching geometries of early vascular plants.

Naturally the three-dimensional universe is an oversimplification of plant geometry. In reality many more factors than the three described here may influence plant shape. It is possible to simulate multidimensional universes of plant shape, incorporating such factors as the ability to change the length of axes or their girth as well as the ability to produce axes with different probabilities of branching. In particular, a more sophisticated simulation might consider a variable that allows for a phenomenon called unequal branching: at each branching point one new axis may deviate more than the other new one does from the orientation of the axis from which they both grow. In this case there are actually two branching angles, each representing the angle between one of the new axes and the original axis. Multidimensional universes, which incorporate such factors, are hard to visualize, but they can easily be simulated with the aid of computers. For purposes of simplification, however, the cubical, or three-dimensional, universe of branching patterns is sufficient to describe the initial phases of plant evolution.

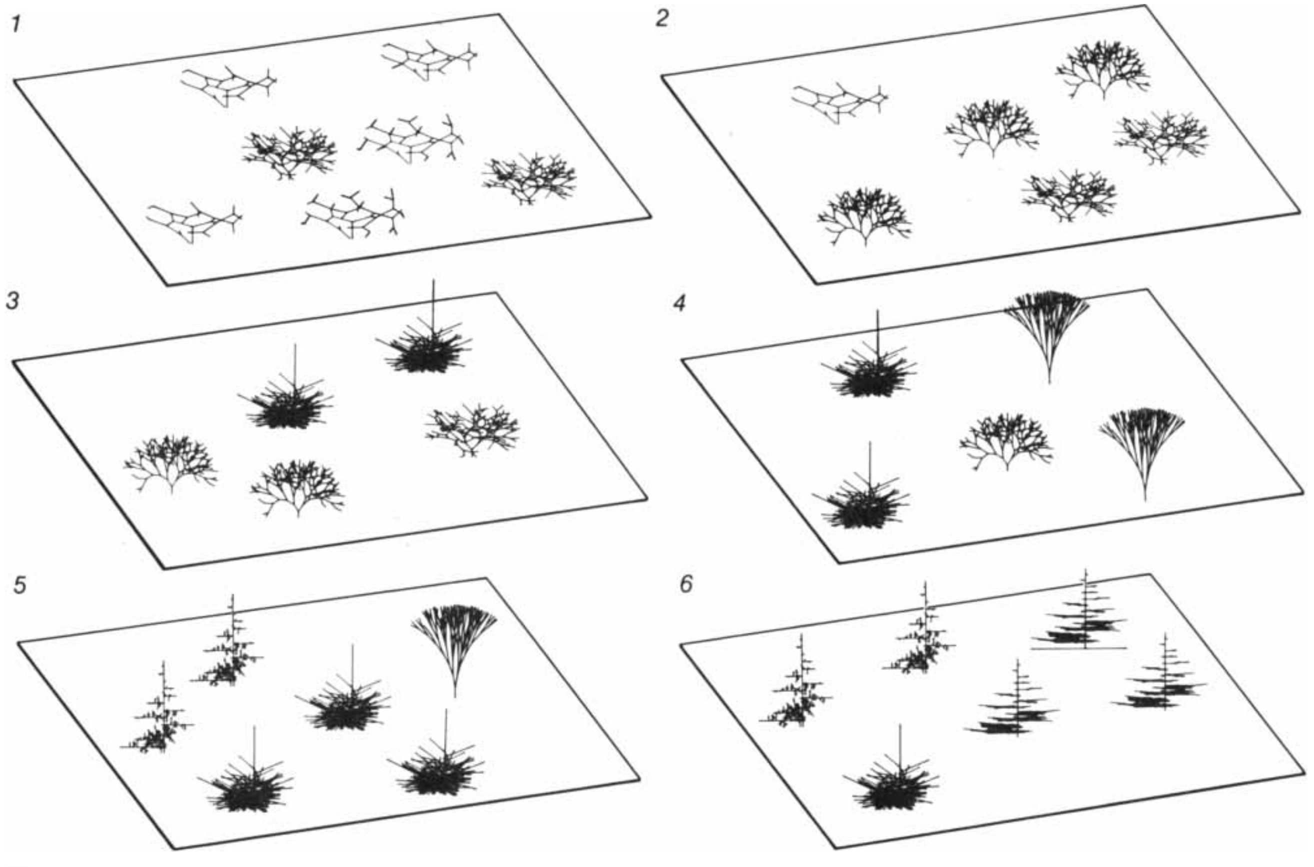
A simulated "evolutionary trajectory" of early land plants can be represented within the cube. First the location within the cube of the most primitive branching pattern (the pattern most closely resembling that of the oldest plants) is determined. Then the computer selects from the plant's nearest neighbors in the cube a species that is more efficient at gathering light. (The light-gathering efficiency of a species is estimated by allowing the computer to grow 10 random samples of the species and find their average light-gathering efficiency.) The process is reiterated until the computer has identified a set of morphological characteristics that is more efficient than any of the species' immediate neighbors in the cube. A line drawn through all points representing the plant species selected by the computer represents an evolutionary trajectory based on competition for light.

It is also possible to produce a trajectory based on balancing the tradeoff between light-gathering efficiency and mechanical stress. In computing such a trajectory the computer, when it scans a plant's near neighbors to determine which species is next in the trajectory, chooses the species that has the highest ratio of light-gathering efficiency to bending moment, rather than simply the plant with the highest efficiency.

The evolutionary trajectories thus simulated, both those based on light-gathering efficiency alone and those designed to reconcile the tradeoff between gathering light and sustaining stress, bear a strong resemblance to trends in the fossil record. The most primitive branching pattern is characterized by few levels of branching, wide angles between axes and nearly vertical planes of branching. The most advanced pattern has planated groups of lateral axes attached to a single main vertical axis.

Intermediate geometries show how the transition from a primitive geometry to an advanced one takes place in the simulated evolutionary trajectory. First the amount of branching increases. Then one axis in each successive branching grows closer to the vertical than the other axes do and so becomes a central axis. Finally, the resulting lateral axes (the ones growing out from the central axis) come to lie on planes that are close to the horizontal. This sequence represents changes that maximize light-gathering efficiency and minimize mechanical stresses. Coincidentally, it turns out to be virtually identical with the sequence that maximizes the amount of shade cast on neighboring plants while minimizing the amount of self-shading.

Although the simulated evolutionary trends match trends found in the fossil record with reasonable accuracy, the correspondence between the simulated and the real trends in plant evolution could be merely coincidental and biologically irrelevant. Other hypotheses might account for the evolutionary record more fully than the notion that plants are driven by pressures to gather light and sustain me-



COMPUTERIZED WAR GAME simulates evolution driven by competition among species. The illustration shows six successive stages in the game; in each stage one or more plants of each of three species compete for light and space. The single species in each stage that gathers the least light is eliminated, and the other two are allowed to disseminate spores. Half of the spores from the most successful species mutate, forming a third species to compete in the next stage of the game. The ability to grow depends on the amount of "light" a plant receives (which in turn depends on the amount of surface it exposes and the degree to which it is shaded by itself and other plants). The number of spores a plant disseminates depends on the number of branch ends it has, and the area over which they are distributed depends on its height. Success in the game combines ability to gather light, to shade neighboring plants and to reproduce.

Here are 128 reasons to buy at Elek-Tek, not to mention the fastest delivery anywhere.

Save 30% to 43%
off Manufacturer Suggested Retail prices on
EPSON • OKIDATA • STAR • TOSHIBA • DIABLO •

HEWLETT-PACKARD

Lower Prices on 4 Hewlett-Packard "Best Seller" Calculators

• HP11C	$43
• HP15C	75
• HP41CV	134
• HP41CX	189

IOMEGA BERNOULLI BOX (Modular)

1.	10 meg ½ height Drive for IBM-PC/XT/AT & compatibles	1675
2.	20 meg ½ height Drive for IBM-PC/XT/AT & compatibles	2335
3.	Non-Bootable Interface Card	104
4.	Bootable Interface Card	234
5.	10 meg cartridges for above (3 pak special)	125
6.	Head Cleaning Kit	69

17.	LX 80	$ 215
18.	RX 100+	299
19.	FX 85	340
20.	FX 286	CALL
21.	LQ 800 parallel	CALL
22.	LQ 1000 parallel	CALL
23.	DX10 Daisy Wheel 10CPS	230
24.	DX20 Daisy Wheel 20CPS	320
25.	DX-35 Daisy Wheel 35CPS	590

Huge Discount on TOSHIBA OKIDATA star

SPECIAL PURCHASE!
P-38 Dot Matrix 400 CPS
Mfr. Sugg. Ret. $1495
Elek-Tek Price $600

SPECIAL PURCHASE!
Diablo D-36 Daisywheel 35 CPS
Mfr. Sugg. Ret. $1495
Elek-Tek Price $450

70.	HP-41CV	$ 134
71.	HP-41CX	189
72.	Optical Wand	95
73.	Card Reader	145
74.	Printer	283
75.	Ext. Memory Module	60
76.	Ext. Function Module	60
77.	Time Module	60
78.	HP-IL Loop Module	95
79.	Digital Cassette Drive	400
80.	Printer/Plotter (HP-IL)	335
81.	HP-9114A Disk Drive	600

HARD DRIVES

Dysan CORPORATION

Internal Hard Disk Subsystems

7.	20 MEGABYTE	500

Each system includes:
- Micro Science Drive
- Adaptek Controller
- 2 Cables
- Documentation
- Dysan 1yr limited warranty
- For use in IBM PC, XT or most compatibles

PRINTER BUFFERS

8.	**DDP-16** 64K Par/Par.	80

QUADRAM Microfazers

9.	**QRMP8** 8K Parallel/Parallel	125
10.	**QRMPS8, SP8 & SS8** Parallel/serial, serial/parallel & serial/serial	139

POWER PROTECTORS

11.	**PowerMite PG600-S9** 6 outlet Surge Suppressor	30

TRIPPLITE

12.	**ISOBAR 4** 4 outlet Surge Suppressor	42
13.	**ISOBAR 8** 8 outlet Surge Suppressor	50
14.	**Line Conditioner** (4 outlet)	135
15.	**Backup Power Supply** (425 watt)	360
16.	**Backup Power Supply** (1000 watt)	850

PRODUCTS FOR IBM-PC

26.	**Amdek 310A** Amber Monitor	$ 150
27.	**Generic Multi** Multifunction Board, 64K	135
28.	**Generic Multi 384K** Multifunction Board, 384K	175
29.	**AST Six Pak +** Multifunction Board, 64K	225
30.	**AST Six Pak + (loaded)** Multifunction Board, 384K	290
31.	**Quadram Quadboard** Multifunction Board, 0K	195
32.	**Quadram Quadboard** Multi. Board, 64K/384K	210/267
33.	**Orchid Tech.** PC Turbo 186	570
34.	ECCELL 0K	399
35.	ECCELL Daughterboard	135
36.	ECCELL I/O Board	99
37.	**Hercules** Monochrome Card	299
38.	**Hercules Color** Color Graphic Card	155
39.	**Novation 490605-1** 2400BPS inc. Mite Software	620
40.	**Novation 490592-1** 1200B Internal Modem No software	150
41.	**Novation 490603** ½ Card Modem 2400BPS No Software	425
42.	**Hayes 1200** External modem	380
43.	**Hayes 1200B** Internal modem w/software	359
44.	**Hayes 2400** External modem	599
45.	**Hayes 2400B** Internal modem w/software	525
46.	**AT&T 4000** 300/1200 Ext. Modem	335
47.	**US Robotics Courier 2400** Ext. 2400B Smart Modem	460
48.	**US Robotics Telpac** Telecomm. Software	75
49.	**Toshiba HND 04DT** ½ ht. DSDD Disk Drive	90

IBM is a trademark of International Business Machines Corp.

PRODUCTS FOR APPLE

50.	**U—Print IIc-16K** Par Interface ext. for IIc	85
51.	**Grappler PLUS** Graphics Interface	80
52.	**Buffered Grappler** 16K Buffer (Expand to 64K)	150
53.	**Kensington System Saver** Fan and Surge Suppr.	69
54.	**Super Cooling Fan** Fan and Surge Suppr.	29
55.	**80 Column Card** for Apple II	55
56.	**Gamma** Disk Controller Card	48
57.	**Gamma II** SS Disk Drive for II, IIe, IIplus	125
58.	**Gamma IIc** SS Disk Drive for IIc	125
59.	**Rana ELITE 2** DSSD Disk Drive	299
60.	**Rana ELITE 3** DSDD Disk Drive	369
61.	**ERAM 80** 80 col; 64K RAM for IIe	115
62.	**Novation Apple Cat II** 300 Baud Modem	200
63.	**Novation 212** 300/1200 Baud Modem	390
64.	**Nov 49059112** Mod. & Soft. for MacIntosh	CALL
65.	**Hayes Micromodem IIe** 300 Baud Modem	145
66.	**Hayes Smartmodem IIc** 300 Baud Modem for IIc	240
67.	**Wico 501030** Analog Joystick	36
68.	**Amdek Monitors**	CALL
69.	**Zenith Monitors**	CALL

Hewlett-Packard QUAD MEMORY MODULE
$25 Mfr. Sugg. Ret. $75
Upgrades the HP41C to the power of the HP41CV

SLIMLINE Shirtpocket Styled Power Packed Programmable LCD PROBLEM SOLVERS

82.	**HP-11C** Scientific	43
83.	**HP-15C** Scientific	75
84.	**HP-12C** Financial	90
85.	**HP-16C** Programmer	90

HP-ThinkJet

HP-2225B $350

86.	**HP-71B Computer**	$ 399
87.	**HP-82400A** Series 70 card reader	125
88.	**HP-82401A** HP-IL interface	95
89.	**HP-82402A** 4K memory module	60
90.	**HP-82700A** 8K memory module	150

PERSONAL COMPUTATION DEALER

HUGE SAVINGS ON DISKETTES

	Dysan	maxell	3M	MEMOREX	BASF
3½" SSDD	20.00	20.00	23.00	—	—
DSDD	27.00	26.00	29.00	—	—
5¼" SSDD	15.00	13.00	12.00	11.50	11.00
DSDD	19.00	15.00	15.00	14.00	12.00
SSDD96TPI	24.00	24.00	24.00	—	—
DSDD96TPI	33.00	29.00	29.00	—	—
5¼" DSDDHD (For IBM AT)	33.00	28.00	30.00	—	24.00
8" SSDD**	22.00	29.00	25.00	—	—
8" DSDD**	26.00	32.00	29.00	—	—

**Unformatted

Call for Quantity pricing for 10 boxes or more.

3M DATA CARTRIDGES

3M		1	10	25
122.	DC100	$13.50	$13.00	$12.00
123.	DC300A	17.50	17.00	16.00
124.	DC300XL	20.50	20.00	19.00
125.	DC300XL/P	21.50	21.00	20.00
126.	DC600A	23.00	22.50	21.50
127.	DC1000	14.50	14.00	13.50
128.	DC2000	19.50	19.00	18.50

CALL TOLL FREE 800-621-1269 EXCEPT Illinois, Alaska **CANADIAN TOLL FREE 800-458-9133**

ELEK-TEK, inc.
6557 N. Lincoln Ave. Chicago, IL 60645
(312) 631-7800 (312) 677-7660

Corp. Accts. Invited. Min. Ord. $15.00. Visa or MasterCard by Mail or Phone. Mail Cashier's Check, Mon. Ord., Personal Check (2 wks. to clear). Add $4.00 1st item. (AK, HI, PR, Canada add $10.00 first item) $1.00 ea. add'l shpg. & handl. Shipments to IL address add 7% tax. Prices subj. to change. WRITE for free catalog. RETURN POLICY: Defectives Only: Most products replaced within 30 days of purchase with identical merchandise only. Computer and large peripherals replaced only when defective on arrival (within 3 work days of delivery). Other problems covered by mfr. warranty. ALL ELEK-TEK MERCHANDISE IS BRAND NEW, FIRST QUALITY AND COMPLETE. Delivery subject to availability. DUNS #09-718-0617 chanical stress. Surely, for example, the success of a species must depend in large part on its success as a reproductive entity. Can the computer examine this hypothesis as well?

To answer the question it is possible to simulate a "war game," in which branching patterns having different light-gathering capacities and reproductive potentials compete with one another in a universe shown on the computer screen. The objective of the game is to declare an "evolutionary winner" on the basis of game rules that are really hypotheses about how evolution might have proceeded.

Each plant is given an area in which it can grow. The area is defined by the plant's ability to cast a shadow over neighboring plants. Each plant is also given an area of a different size, centered on the first, over which it can disseminate reproductive units, which in primitive plants were spores dispersed by the wind. Because many of the early land plants shed spores from the tips of their axes, the number of spores each plant produces is proportional to the number of branch tips in its branching pattern; the area in which the plant may disseminate its spores is determined by its height. (This is aerodynamically reasonable, because the height from which any light or small object is released is critical in determining the distance a breeze will carry it from its point of origin.) The height of a plant depends not only on the number of branchings but also on the plant's branching angle; plants with smaller branching angles grow taller. Any spores that are distributed in a shadowed area, including the one created by their own parent plant, fail to germinate. (This game rule is reasonable too, because many of the lower living vascular plants can inhibit the growth of their own progeny.)

Plants that have overlapping shadow areas interfere with one another's ability to gather light. Each plant's rate of branching, and therefore of growth, is determined by the amount of sunlight it receives. That depends in turn partly on its tendency to avoid self-shading, partly on how heavily it is shaded by other plants and partly on the amount of photosynthetic surface it presents to the sun.

As the game begins, equal numbers of spores of plants having the three most distinct primitive branching patterns are scattered on a level "playing field" and are allowed to grow. After each plant has grown to 10 levels of branching the computer determines the species whose members gathered the least sunlight and eliminates all plants of that species. The other two species are then allowed to disseminate spores. Half of the spores from the species whose members gathered the most sunlight are allowed to mutate: their "genetic" factors are altered slightly in ways that increase the growing plant's ability to cast a shadow, to avoid self-shading, to produce spores or to disseminate spores over a large area. The resulting three branching patterns (the two most successful patterns of the previous round and the new, mutated pattern) are then allowed to grow and the sequence is repeated. The game can be made more complex by introducing mutations that are not advantageous, but this generally modifies the results only a little, and it lengthens considerably the time necessary to finish the game.

The game can take a long time to play, even on a computer, because thousands of calculations must be made for even a brief skirmish. The game ends when the last mutation has been selected from the repertoire offered by the trajectory under study. The winners are the species that survive until the end. The game is thus, to be sure, artificially truncated. Real evolution can produce not only changes in the shape of a plant's branching pattern but also changes in the plant's physiology, such as the ability to tolerate shade or to grow secondary wood. Real evolution can also be punctuated by external events such as natural catastrophes, which restart the game with different proportions of players. Nevertheless, even this simple war game generates trends much like those found in the fossil record, and so the rules probably bear some relation to the biology of real plants.

The trajectory emphasizing light-gathering capacity and the trajectory reconciling tradeoffs between gathering light and sustaining stresses are about equally good at indicating which mutations give rise to plants that do well in the war game. In both schemes the most primitive geometry has only a few levels of branching. With sporangia (spore-bearing structures) borne at the tips of its axes, such a plant will have a small reproductive output. In both schemes the most advanced geometry consists of a single tall axis that bears many levels of lateral axes. Sporangia of such a plant, if borne at the tips of its axes, would be numerous and many would be high above the ground, ensuring dispersal over a wide area.

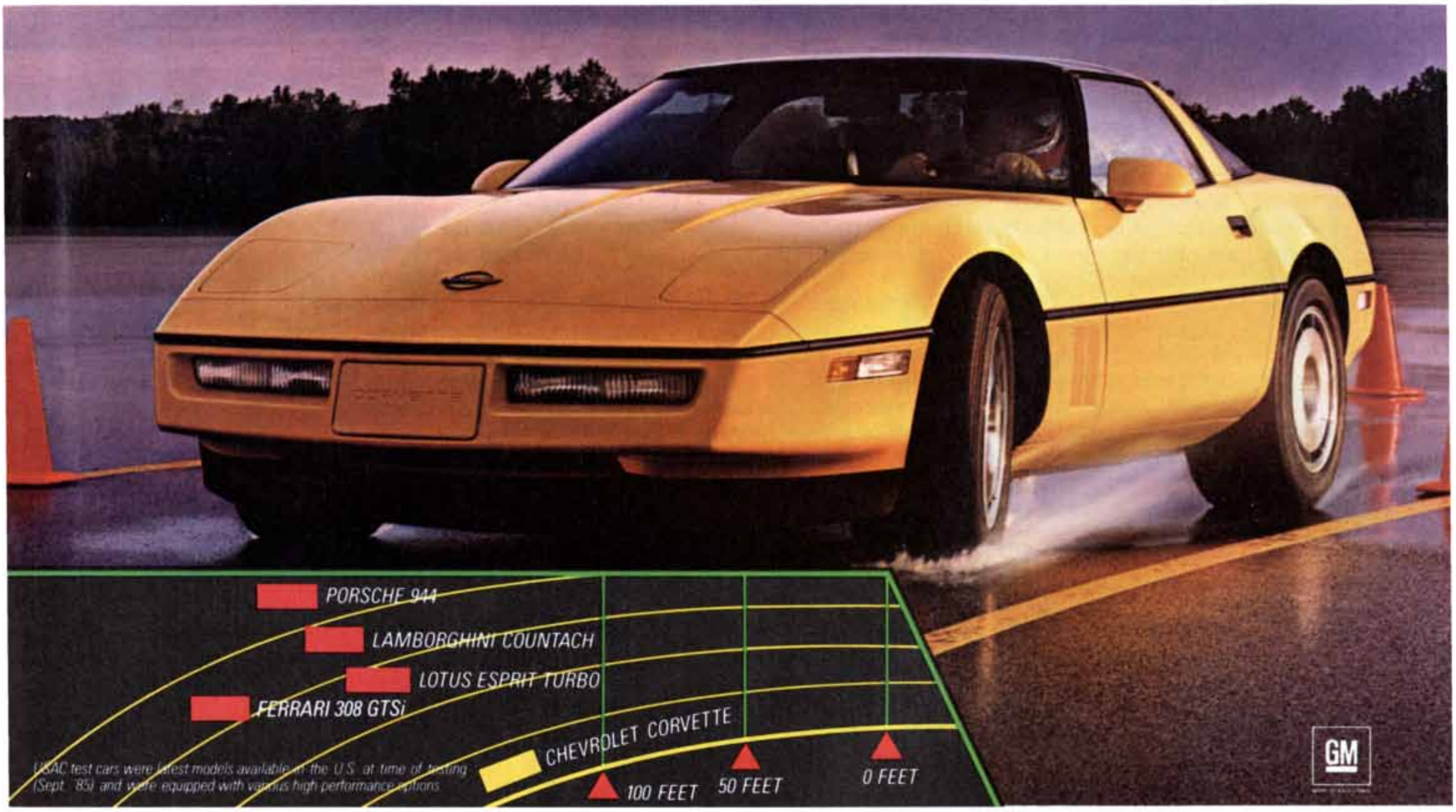
Observations of living plants indicate that the kind of plant that succeeds in the war game is indeed successful in natural competition as well. Studies of the population dynamics of living plant monocultures (cultures of many individuals of the same species) indicate that larger individuals have an advantage over smaller ones. As the population density of the plants increases, the mortality rate increases as well, but a greater proportion of the small plants die than of the large plants. These phenomena have been observed in many species of plants as well as in the war game.

The same pressures that give rise to these tendencies in living monocultures may well have influenced the early evolution of vascular land plants. When plants first invaded the land, the new habitat was probably sparsely occupied, and so the population densities were low. In time both the number of individuals and the number of species at a site would have increased. Species that grew higher or faster would then have had an advantage. As the density of plant communities increased, individuals and even species may have died out selectively, favoring the survival of taller species.

In this context it is interesting to note a seemingly contradictory point. Because of self-shading, added growth tends to reduce light-gathering efficiency: the plant gathers more light, but it gathers less light per axis. This loss of efficiency is counterbalanced, however, by the tall plant's increased tendency to cast shade on neighboring plants and their spores. The trend toward larger plants may therefore have conferred advantages on a species even though it actually impaired the vegetative performance of individual plants. Apparently, then, the mere presence of many individuals of the same species can give rise to evolutionary trends that will help the species to compete against others.

Computer simulations have shown that the various hypotheses proposed concerning plant evolution have consequences that can be verified in the fossil record. One cannot say that the hypotheses have been proved to be correct, only that they have been partially confirmed. Moreover, the simulated consequences of the several hypotheses are virtually identical. It is therefore impossible to determine the relative importance of the various criteria used to produce the simulations; that can be done only when two hypotheses lead to conflicting predictions. Moreover, we have yet to produce simulations modeling such important structural needs as the ability to translocate fluids and to dissipate heat. The general technique of simulation offers great promise as a tool for the evolutionary biologist, but the successful use of any tool requires practice and judgment, and the tool itself must be developed and refined.

Anti-lock braking power you can grade on a curve.



USAC test cars were latest models available in the U.S. at time of testing (Sept. '85) and were equipped with various high performance options.

You're driving 55 MPH on a rain-slick curve. Suddenly the unexpected: You stand on the brake pedal and steer to stay in your lane. You might expect Europe's most exotic cars to handle such a crisis effortlessly. Yet for all its awesome straight-line braking ability, Ferrari 308 GTSi failed to negotiate a 150-foot radius curve at maximum braking in USAC-certified testing. Lamborghini Countach failed. Lotus Esprit Turbo failed. Porsche 944 failed. Only the 1986 Corvette demonstrated the ability to steer and stop in these conditions at the same time. Only Corvette made the turn while coming to a controlled stop. When conditions turn foul, Corvette's new computerized Bosch ABS II anti-lock braking system is designed to help improve a driver's ability to simultaneously brake and steer out of trouble.

Why does the Corvette feature the world's most advanced braking technology? Because a world-class champion should give you the edge in an emergency.

Corvette. A world-class champion.

TODAY'S CHEVROLET Live it!

Mental Imagery and the Visual System

What is the relation between mental imagery and visual perception? Recent work suggests the two share many of the same neural processes in the human visual system

by Ronald A. Finke

People often report that they can form mental images of an object that resemble the object's actual appearance. The act of constructing such images often produces visual sensations that seem quite realistic. Imagine, for example, that you are looking at an elephant. Does it have a curved trunk? What color are its tusks? How big are its eyes? Most people contend they attempt to answer such questions by "inspecting" a mental image in much the same way as they would inspect a real elephant.

These informal observations about imagery naturally lead one to consider the extent to which imagery and visual perception might be related. They suggest in particular that mental imagery may involve many of the same kinds of internal neural processes that underlie visual perception, a possibility that would have important theoretical and practical implications. If it could be established, for instance, that mental imagery shares with visual perception common neural mechanisms in the human visual system, one could begin to establish just how imagery may interact with visual perception. This would make it possible to explore the various ways imagery could function to facilitate, enhance or even substitute for visual perception.

For the past 10 years my colleagues and I have been developing techniques for investigating the functional relation between mental imagery and visual perception. Because experimental subjects can often guess what ought to happen in an imagery experiment, we have striven to make our techniques precise enough to reveal subtle correspondences between imagery and perception. Our work has revealed that mental images display a much richer variety of visual properties than had been previously thought, but also that imagery differs from perception in certain respects.

Through introspection one can recognize that features of a mental image formed at a small size or a far distance are harder to distinguish than features of an image formed at a large size or a near distance. Try, for example, to imagine an ant on a newspaper several feet away and then on the tip of a toothpick directly in front of your eyes. You should be able to mentally "see" many more of the ant's features (such as its head and body segments) when you imagine it at close range.

Stephen M. Kosslyn of Harvard University explored this relation between image size and feature resolution by employing simple reaction-time techniques. He found that the features of an imagined animal, such as the eyes and ears of a cat, could be detected more quickly when subjects were instructed to fashion relatively large images or assume a relatively close vantage. The experiments were inspired by the common observation that features of real physical objects can be detected faster when they are viewed from a closer distance.

More recently Howard S. Kurtzman and I have done experiments at Cornell University to measure precisely how well the features of objects can be resolved, or distinguished, in imagery and in perception. We were particularly interested in how the size of the features, their spacing and their position in the visual field affected resolution, or the ability to distinguish among details. We predicted that across all these variations visual resolution in mental imagery should match the resolution in perception.

Resolution in visual perception falls off continually as one observes an object at locations progressively farther from the point of eye fixation. The amount of detail that can be distinguished is not the same in all directions, however. As a rule resolution decreases more slowly along the horizontal axis of the visual field than along the vertical axis, and more slowly below the point of fixation than above it. It is also known that bar gratings become harder to resolve as the gratings become increasingly finer—more precisely, as their fundamental spatial frequency increases.

Our method for measuring limits of resolution in mental imagery was based on certain techniques common to visual psychophysics. Initially we showed our subjects a flat disk whose upper half was filled with a series of vertical bars and whose lower half was filled with a series of horizontal bars. The bars in both gratings were the same width. We then instructed our subjects to form a mental image of the disk and to project the image on the center of a screen directly in front of them. On the screen eight lines extended radially out from the center. The subjects indicated how far they could

WATERCOLOR LANDSCAPE was painted by a blind Scotswoman who works from her mental images. The artist, Carolyn James, suffers from a particularly acute form of the eye disease known as retinitis pigmentosa. Now 42 years old, she was registered blind at 21. To paint she lines up 24 watercolor jars in front of her in a memorized order. She moves from section to section on the paper, determining what she has just finished by detecting the moisture with her fingertips. Each of her watercolors is typically composed of six layers of paint.

look away from their images along each of the lines on the display before they could no longer tell the two halves of the imagined pattern apart. They reported that the gratings appeared fuzzy and indistinct as they were imagined farther into the visual periphery and then could no longer be distinguished beyond a certain point. For comparison, the same judgments were also obtained in a perception condition, in which the same disk was actually projected on the center of the screen in front of the subjects.

We repeated the experiment for each of three disks. The bar widths of the second disk were three times thinner than those of the first; those of the third were three times thinner than those of the second. On the average the fields of visual resolution decreased in size with increasing spatial frequency (decreasing bar width), and they were virtually identical whether the gratings had been imagined or observed. The imagery and the perceptual fields were also very similar in shape: resolution decreased more slowly along the horizontal axis than along the vertical axis, as well as more slowly below the point of fixation than above it.

We then did a control experiment in which we showed a different group of subjects the original set of three disks and asked them to predict the image-resolution fields. Their predictions differed considerably from what we had observed, arguing against a trivial "guessing" explanation of our original findings.

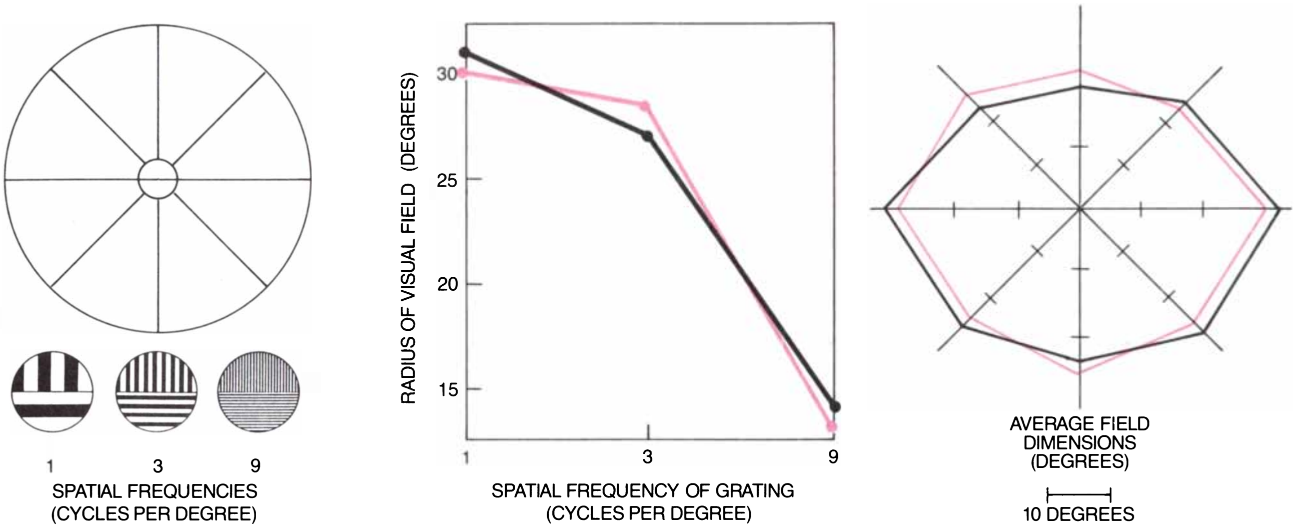
We interpret the results as evidence that pattern discrimination in imagery is constrained in much the same way that it is in visual perception. We propose in addition that these mutual constraints are probably imposed at the pattern-processing levels of the visual system, where the properties of certain neural mechanisms may limit the ability to resolve small or narrowly spaced visual features. Kurtzman and I have done other experiments that are consistent with this result. We found no correspondence between judgments of images and judgments of objects involving differing amounts of visual contrast, or relative brightness, among features. These aspects of perception are thought to be constrained by more primitive kinds of neural processes operating below the levels at which pattern processing takes place.

Our image-resolution findings were based on mental images of flat, two-dimensional patterns. Mental imagery is typically three-dimensional, however; it depicts how objects look in depth as they are viewed from various vantages. When most people imagine their living room, for instance, they can mentally "see" that certain pieces of furniture are in front of others, depending on where in the living room they imagine themselves to be.

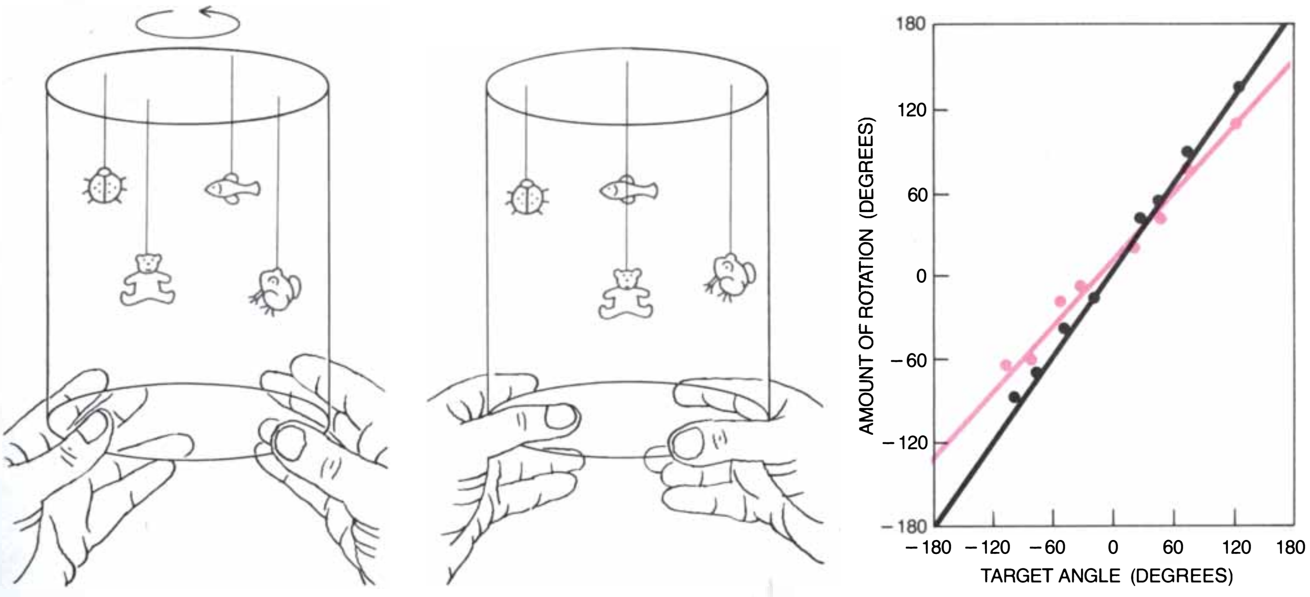
To investigate the three-dimensional properties of images, Steven Pinker and I asked subjects to form and mentally rotate images of a configuration of objects in space. When one actually looks at a three-dimensional configuration of objects from different perspectives, the objects are usually seen to shift their relative position in depth as the viewer moves or as the configuration is rotated. Recall, for example, times when you may have watched

people riding on a merry-go-round. You probably noted that the people would appear to shift their locations in relation to your vantage as the merry-go-round turned, perhaps forming familiar two-dimensional patterns at certain moments in much the same way that a constellation of stars often forms flat, recognizable patterns. In our experiments, which we did at Harvard, we were particularly interested in finding out whether similar kinds of patterns would appear to emerge when subjects imagined looking at a rotated configuration of objects. The results indicate striking similarities.

We asked our subjects first to learn the locations of four small plastic animals suspended at different heights in a transparent cylinder, and then to form mental images of each of them after



CONSTRAINTS on visual resolution, or the ability to distinguish details, were measured by means of the three patterns shown at the bottom left. Experimental subjects were instructed to form mental images of each pattern and to project their images onto the center of a large circular display (*top left*). The subjects then indicated how far they could look away from their images along each of the eight lines on the display before they could no longer tell the two halves of the imagined patterns apart. The visual fields within which the bar gratings of the imagined patterns could be distinguished decreased in size with increasing spatial frequency, or decreasing bar width (*colored line in middle*). On the average these fields were elongated horizontally and were larger below the direction of gaze than above it (*colored shape at right*). Virtually identical results (*black line in middle and black shape at right*) were obtained when the patterns were actually projected on the display, indicating that similar constraints are imposed, on feature resolution in imagery and in perception.



THREE-DIMENSIONAL PROPERTIES of mental images were explored with a transparent cylinder. Subjects were told to learn the locations of four small plastic animals suspended at different heights in the cylinder and then, after the animals had been removed, to form mental images of them. As the empty cylinder was rotated 90 degrees, the subjects rotated their mental images (*left*). Although the appearance of the objects in the original viewing position did not suggest that a parallelogram with each animal at a corner would emerge, the subjects detected such a pattern. Their drawings of the pattern, however, showed small but systematic distortions. An explanation for the occurrence of these distortions was suggested by another experiment, in which the subjects rotated the cylinder manually in order to align pairs of the imagined animals vertically (*middle*). To align the imagined animals, subjects rotated the cylinder by a consistently smaller amount (*colored line at right*) than they did in a similar test with the animals physically present (*black line at right*). In other words, the subjects had mentally rotated their images ahead of their manual rotation of the cylinder.

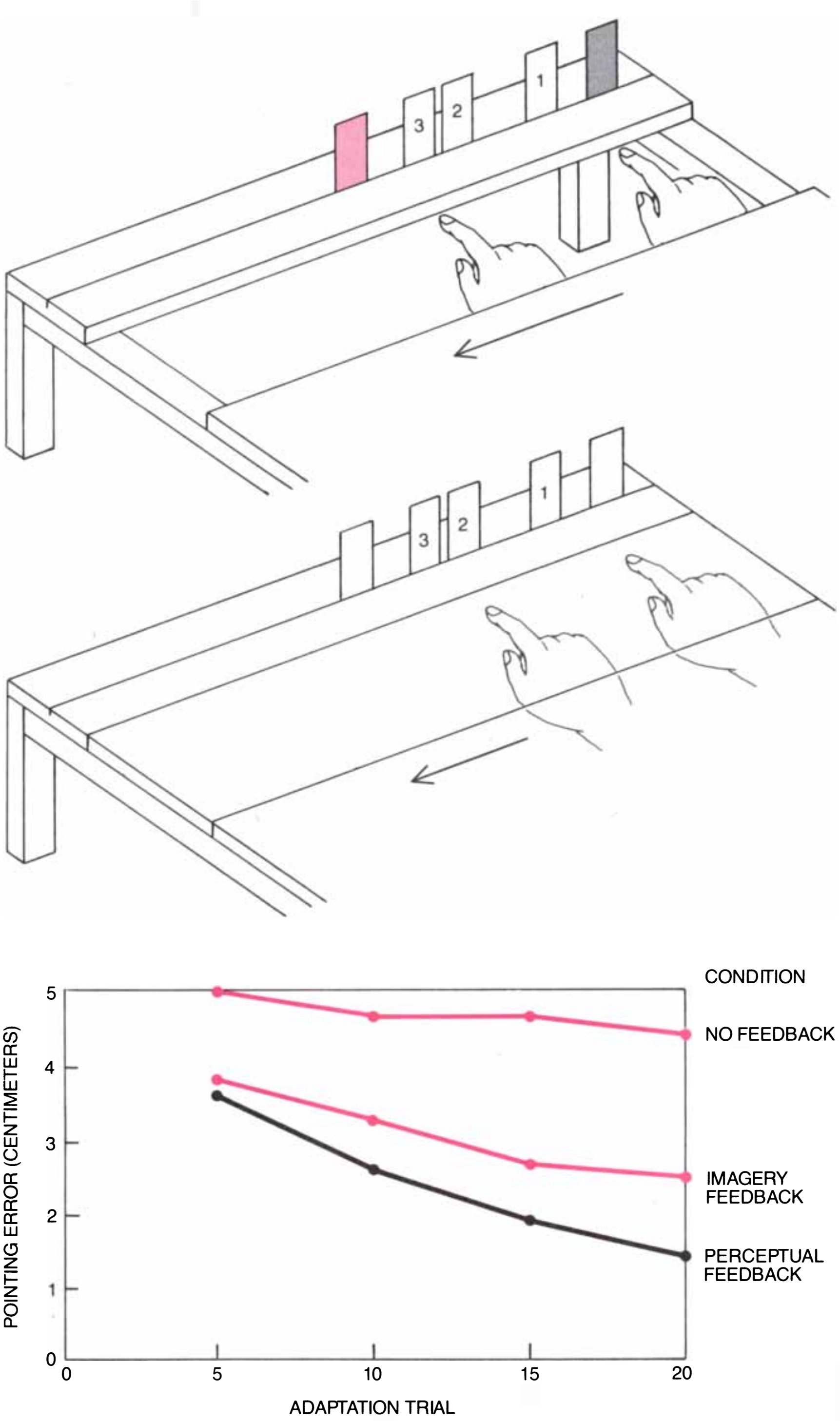
they were removed. We next rotated the empty cylinder 90 degrees and instructed the subjects to draw the imagined configuration as it now appeared from the new vantage. If they had imagined and rotated the animals with perfect accuracy as we rotated the cylinder, the animals would have seemed to form a parallelogram.

In every case the subjects' drawings revealed that their rotated mental images had depicted a pattern closely resembling the parallelogram, even though the appearance of the objects from the original viewing direction did not suggest that this particular geometric form would emerge. Curiously, there were small but systematic distortions in the drawings. The nature of the distortions suggested that the mental rotations had been less than perfectly accurate.

An explanation of these small distortions was suggested by another experiment, in which the subjects manually rotated the cylinder in order to align pairs of the imagined animals vertically. To our surprise we found that the subjects aligned the imagined animals by consistently rotating the cylinder less than was necessary when the animals were physically present. In other words, the subjects had mentally rotated their images ahead of their manual rotation of the cylinder. This tendency to advance an image ahead by small amounts accounted for the minor distortions we had found in our subjects' drawings of the emergent patterns. The experiment thus strengthened our contention that people can accurately imagine the visual perspectives offered by three-dimensional displays. Moreover, it enabled us to measure properties of mental images that naive subjects would not ordinarily expect.

Showing that subjects cannot guess the outcome of an imagery experiment does not, however, rule out the possibility that their performance could be based on unconscious knowledge about changes in the visual appearance of objects—knowledge that could indirectly influence judgments about images. One method for addressing this problem is to have subjects imagine events so atypical or unnatural that the events could not have been previously experienced. If under these conditions behavioral responses obtained from imagery still correspond to those obtained from perception, the imagery performance could not be attributed to the influence of earlier perceptual experiences.

In a series of experiments I carried out at the Massachusetts Institute of



FUNCTIONAL VALUE of imagery was assessed by considering the role images might play in prism adaptation. Optical prisms displace the apparent location of an object. Subjects wearing special glasses containing such prisms were instructed to point to a target (*colored marker in top.row*). As a result of the effect of the prisms they at first pointed about five centimeters to the right of the marker (*gray marker*). Since the prisms displace everything in the field of view, however, once the individuals had extended their arms they could see their error and correct for it in successive attempts (*markers 1–3*). The markers were then utilized by a second group of subjects known as the imagery subjects. They too wore the special glasses, but the space between them and the table holding the markers was covered by a board so that they could not see the fingers of their extended arms (*second row*). Their task was to imagine that they saw their pointing finger arrive under the appropriate error marker as soon as their arms were fully extended. Subjects in a third group, the control group, pointed to the colored marker without being able to observe their errors or being told to imagine them. The graph shows that a significant amount of error reduction ensued when the subjects imagined their errors—almost as much as when they actually saw the errors.

Technology, I attempted to provide this kind of evidence for the functional value of imagery by considering the possible role images might play in prism adaptation. Optical prisms displace the apparent location of objects. A large body of research has shown that people quickly adapt to observing the world through such prisms provided they can move about and note their errors. When the prisms are removed following adaptation, people proceed to make errors of movement in the opposite direction, reflecting changes in their visual-motor coordination. My experiments demonstrated that prism adaptation can occur even when people point at a target and merely imagine they are making errors of movement like those typically induced by displacement prisms.

The subjects in my experiments wore special glasses containing optical prisms. I asked the subjects in one group to extend their right arms and point to a red marker positioned at eye level on a table in front of them. Owing to the effect of the prisms they at first pointed about five centimeters to the right of the marker. Since the prisms displace everything in the field of view, once the individuals had extended their arms they could see their errors and correct for them during a series of subsequent attempts.

I measured and averaged the errors over consecutive groups of trials and then displayed the average error locations with three markers. The markers were for the use of a second group of subjects known as the imagery subjects. These subjects also wore the special glasses, but the area between them and the table holding the markers was covered by a board so that they could not see the fingers of their outstretched arms. I instructed the imagery subjects to point to the red marker while looking through the prisms and then to imagine they saw their pointing finger arrive under the appropriate error marker as soon as their arm was fully extended.

The error markers, in other words, ensured that the imagined errors would correspond to the average pointing errors made by the first group of subjects. I also included a third group of subjects as controls. The individuals in the control group pointed to the red marker without having the benefit of either observing their errors or being told to imagine them.

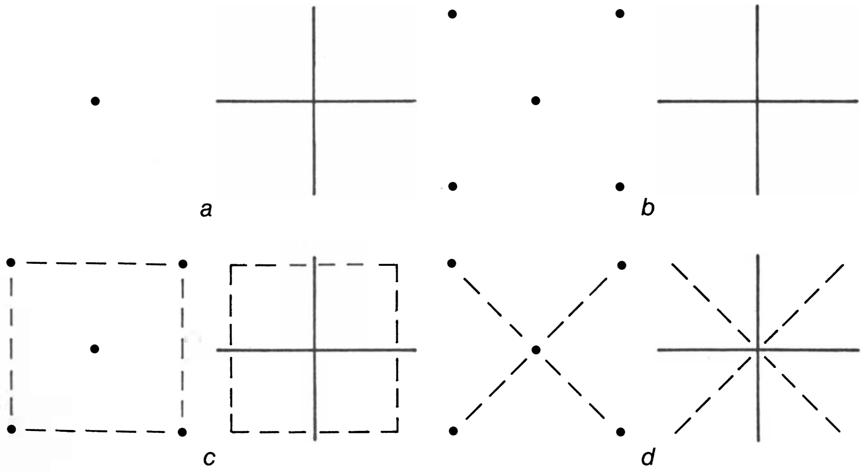
Only the subjects in the perception and imagery conditions showed a significant reduction in pointing error. Moreover, their rates of adaptation were similar. The results for pointing aftereffects, which take place when the glasses are removed, provide additional support for a functional equivalence between observed and imagined errors. Although the aftereffects in the imagery condition were smaller than those in the perception condition, the subjects in both groups pointed to the left of the red marker when normal viewing conditions were restored. I also found evidence of intermanual transfer: the subjects not only pointed to the left with their right hand (the "adapted" hand) but also pointed to the left with their left hand (the "unadapted" hand).

The findings have several implications. First, it is highly improbable that the subjects could have predicted the adaptation and transfer characteristics of prism-induced changes in their visual-motor coordination. It also seems unlikely that they could have had any related kinds of visual experiences providing them with unconscious knowledge of such effects. Second, the findings show that mental imagery can produce certain changes in visual-motor coordination that persist even after the images are no longer formed. They also suggest that the utilization of mental imagery to precipitate such changes may have important practical applications. Professional athletes, for example, often report they find it helpful to rehearse their performance mentally; in the light of these experiments it is reasonable to expect that the success of such techniques depends on the clarity and accuracy with which the performance is imagined.

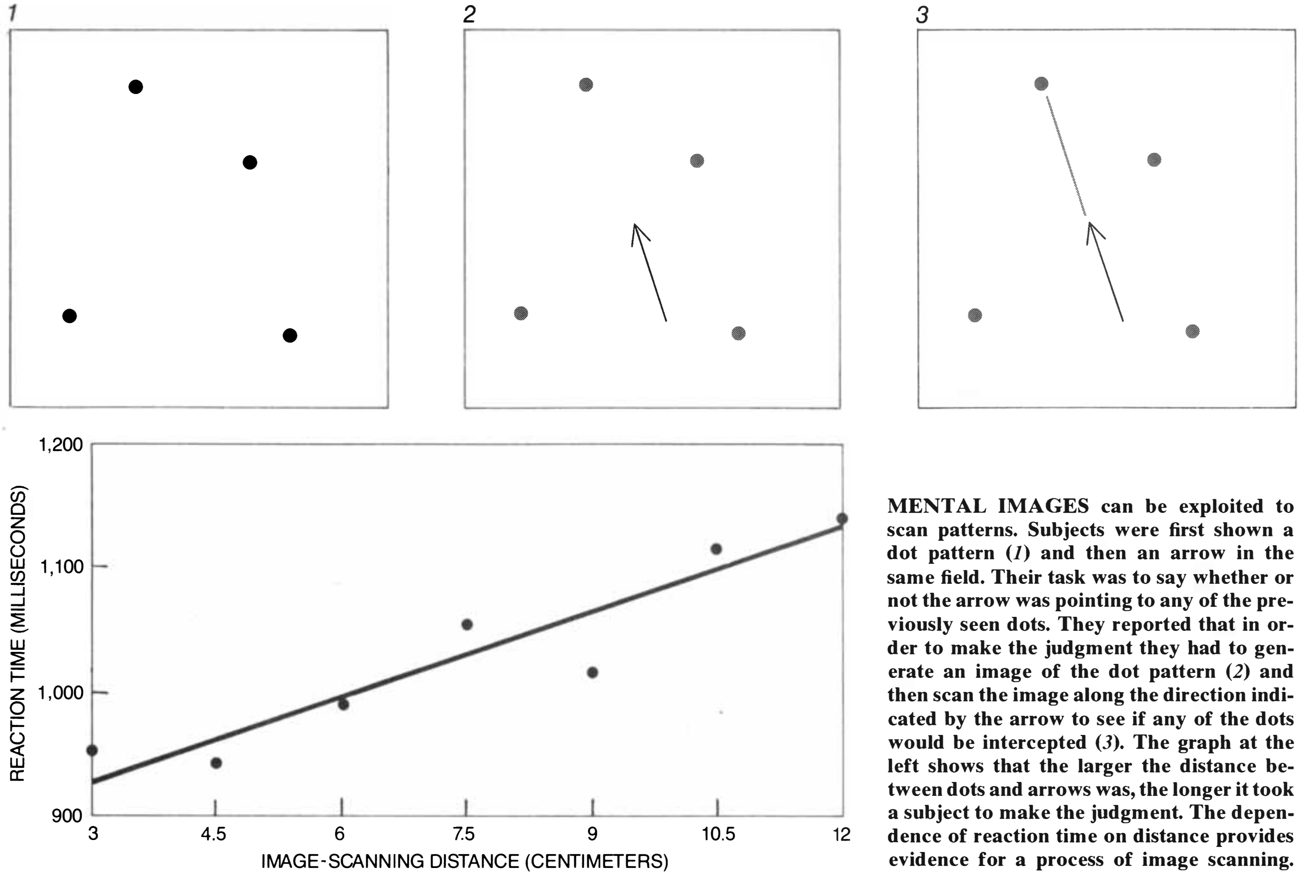
The research findings I have discussed so far illustrate many ways mental images can correspond functionally to physically perceived objects and events. A question of greater practical significance is whether mental imagery can directly facilitate ongoing perceptual processes, assuming that a functional equivalence occurs. Given that similar constraints are imposed on visual resolution in imagery and perception, would it be possible, for instance, to see an object more quickly if an appropriate mental image of the object were formed in advance of its actual appearance?

As proposed some 10 years ago by Ulric Neisser and Lynn A. Cooper, then at Cornell, and Roger N. Shepard of Stanford University, the process of forming a mental image can serve a perceptual anticipatory function: it can prepare a person to receive information about imagined objects. Mental imagery may therefore enhance the perception of an object by causing the selective priming of appropriate neural mechanisms in the visual system. In other words, forming a mental image of an object might initiate certain neural events that are equivalent to those occurring at the moment the object is seen, thereby facilitating the perceptual process.

If an object appears that is different from the one imagined, however, the formation of the image might interfere with the normal operation of the visual system. Suppose, for instance, that you are flying an airplane through



DISCRIMINATION TASK was assisted by mental imagery. Subjects were asked to determine whether a horizontal or a vertical line was the longer. In condition *a* the subjects looked at a fixation point and were then shown the two lines centered at that point. In *b* they were first shown four dots surrounding the fixation point. In *c* they were shown the same dots as in *b* and were asked to form a mental image of a square frame connecting them. In *d* they were asked to form an image of an × through the four dots. The greatest facilitation in the length-discrimination judgments came from imagining the square frame in advance.



MENTAL IMAGES can be exploited to scan patterns. Subjects were first shown a dot pattern (*1*) and then an arrow in the same field. Their task was to say whether or not the arrow was pointing to any of the previously seen dots. They reported that in order to make the judgment they had to generate an image of the dot pattern (*2*) and then scan the image along the direction indicated by the arrow to see if any of the dots would be intercepted (*3*). The graph at the left shows that the larger the distance between dots and arrows was, the longer it took a subject to make the judgment. The dependence of reaction time on distance provides evidence for a process of image scanning.

a cloud. You might see the runway sooner if you were to imagine it in advance at its proper location. If, on the other hand, you imagined the runway at a different location, you might take longer to see it correctly than if you had not imagined it at all.

Some recent experiments I did at Stanford help to clarify the practical relation between image formation and object perception. In one of the experiments subjects indicated whether they saw either a horizontal bar or a vertical bar on a circular screen. These two alternative bars were known as target bars. I told the subjects to form, in advance, a mental image of an identically shaped bar oriented somewhere between the horizontal and the vertical, or to form no mental image. In each trial, therefore, one of the two alternative target bars was superposed on an imagined bar (in those cases where a bar was imagined). I recorded the reaction time for identifying the target bar as a function of the relative alignment between it and the imagined bar. The reaction times for the no-image trials served as a baseline measure.

The subjects made the quickest bar identifications when the imagined bars were closely aligned with the target bars, within a range of about 10 degrees. As the angle between the imagined and the target bars was increased to 45 degrees the identification time also increased. For angles greater than 45 degrees the time decreased once again. In other words, the maximum interference in identification took place when the bars had been imagined to lie exactly in between the two possible target orientations.

Why did the reaction times simply not increase in direct proportion to the degree of misalignment between the imagined and the target bars? One reason is that the subjects' selection of responses could have been based on a comparison between the mental image and the target. If the image matches the target, the response corresponding to the image orientation is quickly selected. If the image and the target differ by 90 degrees, the comparison indicates that the response opposite to the one corresponding to the image orientation is correct. If the imagined bar is in between the two target orientations, however, the comparison becomes confusing and the image interferes with the decision process.

A second experiment supports this explanation. In it I instructed the subjects to indicate as quickly as possible whether either of the two target bars appeared. In the previous experiment they had to distinguish between the two targets; in this case they only had to detect the presence of any target, without having to identify it. The results of this experiment clearly show that mental imagery did not affect simple detection judgments under those conditions. It seems, therefore, that even though image formation may influence the identification of visual patterns, it may not influence the more elementary process of simply detecting any stimulus change.

Additional experiments that Jennifer J. Freyd and I carried out at Stanford provide evidence for a kind of image facilitation that cannot be explained on the basis of response selection. In these experiments we studied the effects of forming a mental image that could serve as a helpful or unhelpful visual context for making difficult length discriminations. We presented our subjects with patterns consisting of two straight lines that formed a simple cross and asked them to indicate which line was longer. At the beginning of certain trials we told the subjects to form an image of an outlined square, which if actually superposed on the center of the line pattern would have enhanced the small differences in the lengths of the lines. During other trials we told the subjects to form an image of an × (the endpoints of which corresponded to the four corners of the imagined square), which would presumably not have been as useful for making the discriminations.

We found that forming a mental image of the helpful context pattern (the square) reduced the time needed to make the length discriminations compared with the time required when the unhelpful pattern (the ×) was imagined or when subjects were not told to form images. Moreover, the effects were similar to those obtained when the same context patterns were actually presented.

We also found that subjects often chose to imagine the helpful context pattern when they were presented with positional cues they could use to imagine either pattern. Since the context patterns themselves could not have biased the selection of the two response alternatives, this type of image facilitation could not have resulted from an internal matching or response-selection process. Instead it may be due to a mental synthesis of real and imagined features at some higher level of the visual system—where the addition of context information can enhance differences among objects that are being compared.

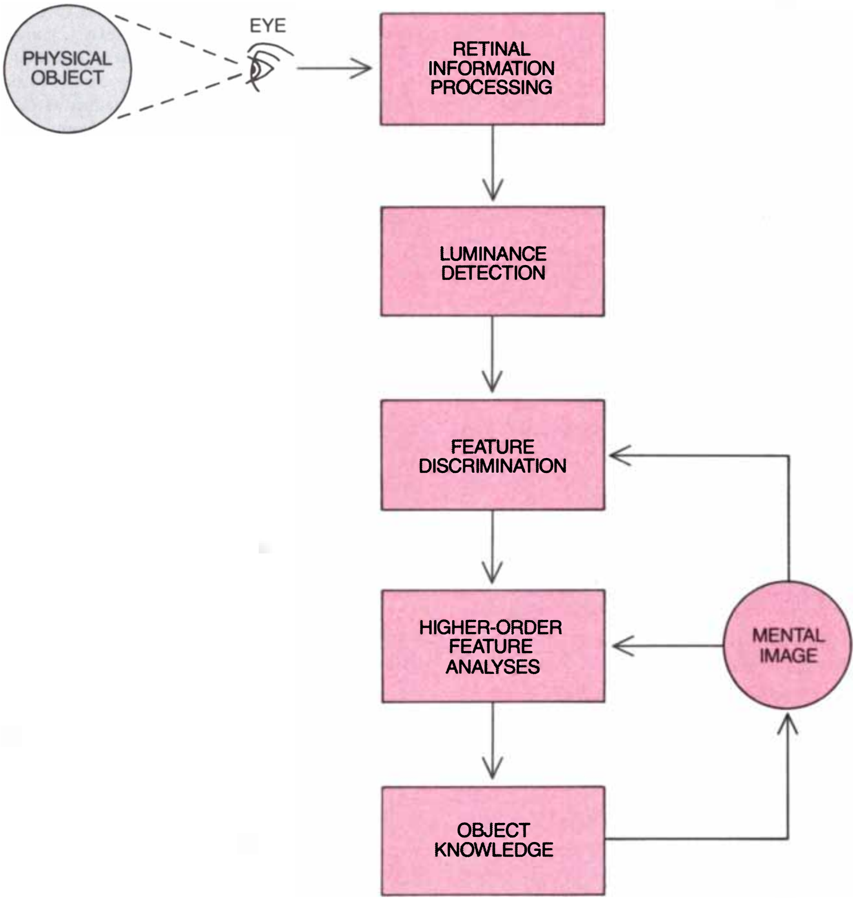
In each of the techniques described up to this point experimental subjects were told explicitly to form some kind of mental image. A possible difficulty with this procedure is that it may encourage the subjects to try to perform as they would in a corresponding perceptual task, thinking that is what they are supposed to do. Although the problem can be largely avoided by attempting to measure subtle or unexpected perceptual effects, an even better way is to show that images can be formed spontaneously for some specific purpose even when no imagery instructions of any kind are given.

The importance of these considerations follows from early studies done by Kosslyn and Pinker on mental-image scanning. They asked subjects to inspect a configuration of objects (such as landmark items drawn on a map), form a mental image of the configuration and "focus in" on one of the objects. The investigators then named a second object and told the subjects to mentally scan along a direct path from the first object to the second. Kosslyn and Pinker consistently found that the time required to complete the mental-image scanning was directly proportional to the original physical distance between the objects, and they therefore concluded that mental images preserve the spatial characteristics of a physical display.

Their findings have been criticized because it would not be hard for experimental subjects to figure out that greater distances should require longer scanning times. Pinker and I have since developed a task that seems to avoid this problem by requiring a subject to form mental images and scan them without explicit directions to do so. After they had inspected a dot pattern, our subjects were shown an arrow and were asked to indicate whether it pointed at any of the previously seen dots. We had predicted that, in order to see if any of the dots would be intercepted, the subjects would have to scan a mental image of the pattern along the direction specified by the arrow.

The experiment turned out to be successful. The decision times increased linearly as distance along the scan path between the arrows and the dots increased. Moreover, nearly all the subjects reported that in order to perform the task they had to form and scan a mental image of the dot pattern. We thus showed that mental-image scanning can be useful whenever it is necessary to anticipate the consequences of moving along a particular path from a given starting point.

Suppose you were trying to figure out where a billiard ball would come to rest on a billiard table after you had aimed it in a certain direction. Even if you could not actually roll the ball across the table or determine the answer mathematically, you could still imagine what would happen by mentally following the motion of the ball and its reflections off the cushions. Cooper and Shepard have reported



MODEL of how mental imagery may influence visual perception is based on the work of the author and other investigators. The perception of an object is a consequence of neural activation within a sequence of information-processing stages in the visual system, beginning at the retinal level. The formation of an image of the object is determined by the knowledge a person has about the features of the object, and presumably it occurs at the very highest levels. Once formed, an image may affect neural mechanisms, at intermediate visual levels, that are responsible for feature discrimination and other more complex types of analyses, thus perhaps modifying perception of the object. Mental imagery probably does not influence visual levels below those concerned with feature discrimination, however.

related findings for the imagined consequences of rotating objects [see "Turning Something Over in the Mind," by Lynn A. Cooper and Roger N. Shepard; SCIENTIFIC AMERICAN, December, 1984].

In the light of these studies it seems reasonable to propose that whenever imagery and perception share common neural mechanisms in the visual system, imagery could facilitate the perceptual processes those mechanisms support. One should therefore seek to determine the lowest visual levels at which such mechanisms may be shared. If visual pattern perception, for instance, is conceived of as involving an orderly sequence of information-processing stages ranging from the lowest to the highest levels of the visual system, one might begin by trying to discover how far down in this sequence image formation can influence the underlying mechanisms.

Starting at the very lowest, or retinal, level, where the most primitive types of information-processing mechanisms are found, one would not expect mental imagery to have much effect. Nor would mental imagery be expected to alter information processing at precortical levels, where mechanisms are responsible for detecting changes in brightness or contrast. Only at the somewhat higher levels responsible for pattern discrimination (as in the visual cortex) does one begin to find evidence that mental images can influence perception. At still higher levels the evidence is strong that imagery can influence perception.

Finally, at the very highest levels one may assume that perceptual processes interact with more abstract processes having to do with knowledge about and understanding of physical objects. Here it is helpful to make a distinction between the form and the function of a mental image. When a person decides to create a mental image of a particular object, the kind of image that can be fashioned depends on the knowledge the person has about the object, such as its size, color and shape. Then once the image is formed it can begin to function in some respects like the object itself, bringing about the activation of certain types of neural mechanisms at lower levels in the visual system. Accordingly, whatever constraints such mechanisms put on the quality of one's perception of the object are also placed on the quality of one's mental imagery. In this way mental images may come to acquire visual characteristics and may serve to modify perception itself.

Crop Storage at Assiros

At a site in northern Greece charred fragments of grain from crop storerooms that burned to the ground 3,000 years ago are throwing new light on how the majestic Bronze Age Mycenaean palaces arose

by Glynis Jones, Kenneth Wardle, Paul Halstead and Diana Wardle

The Bronze Age societies of the Aegean, including the Minoan culture of Crete and the Mycenaean culture of the Greek mainland, were rich and complex. At their center was the palace, which was the seat of government and the focus of religious ritual. The palace also served another crucial function: it was the place where agricultural surplus was stored. Indeed, the authority of Bronze Age rulers in the Aegean rested in large measure on their ability to command supplies of wine, grain, olive oil and other products from the surrounding region. The economic and social arrangements underlying such an elaborate system of food storage must have taken a long time to develop, and yet their history is not easily traced. The primary evidence—stored crops—is eminently perishable. Furthermore, objects of clay, metal or stone have traditionally held greater interest than plant remains for archaeologists studying the Bronze Age. Consequently the patterns of food storage that led up to the palace economy are not fully understood.

Recently, however, the situation has begun to change as students of the Aegean Bronze Age have acquired a keener appreciation of what can be learned from the botanical record. One result of the recent trend is that when we excavated a Bronze Age village at Assiros in northern Greece, beginning in 1975, our ranks included an archaeobotanist (Jones) throughout the dig. Her participation was amply justified by what we found. Careful sampling of seeds from the floors of buildings that had been destroyed by fire during the Bronze Age showed that some of those buildings were specifically devoted to storing crops. Crop storage at Assiros appears to have been communal, taking a form intermediate between storage by single families and the organized appropriation of surplus that was carried out in the mature palace economy. Assiros may be the type of community that would, under the right conditions, have developed into a palatial center. Hence the excavation there may ultimately shed considerable new light on the origins of the great Bronze Age palaces.

Assiros, the site that has provided these new insights, is in Macedonia. Today Macedonia is part of Greece, but during the Bronze Age it lay on the fringe of the Mycenaean world. Mycenaean society was centered on the Aegean peninsula; its northern limit ran south of Macedonia through Thessaly. In the Bronze Age Macedonia was a cultural crossroads, subject to influences from the Danube region to the north, from Thrace to the east and from Albania to the west as well as from Mycenaean society. Among the varied influences, that of Mycenaean society was one of the strongest. It is clear, however, that the social development of the northerners was less advanced than that of their southern neighbors at Mycenae and Pylos on the mainland and at Knossos on Crete. Nothing found in Macedonia can be compared to the great palaces that have been uncovered at those sites and at others [see "Minoan Palaces," by Peter M. Warren; SCIENTIFIC AMERICAN, July, 1985]. Such relative backwardness can be quite helpful to archaeologists, because it implies that findings from Assiros that are contemporary with the palaces can provide information about the preconditions of the palace economy.

Indeed, one reason Assiros is such a good potential source of information about the antecedents of the palace economy is that the period of occupation there neatly straddles the Mycenaean age of palaces. The era of palaces reached its height between 1400 and 1200 B.C., and our work at Assiros shows that the site was continuously inhabited for a millennium spanning those dates. The Bronze Age site we excavated is an oval mound called Assiros Toumba not far from the modern Greek village of Assiros. The mound is about 14 meters high, and our excavation revealed that it is entirely artificial. The first inhabitants arrived in about 1800 B.C. and built a defensive circuit wall enclosing dwellings on what was then a more or less level plain. During the next 1,000 years the circuit wall and the structures within it were rebuilt many times. With each rebuilding the mound rose, and because its sides sloped inward, the habitable area inside the wall decreased. By about 800 B.C., early in the Iron Age, the site was abandoned, perhaps because the enclosed space was no longer large enough to accommodate the population of the settlement.

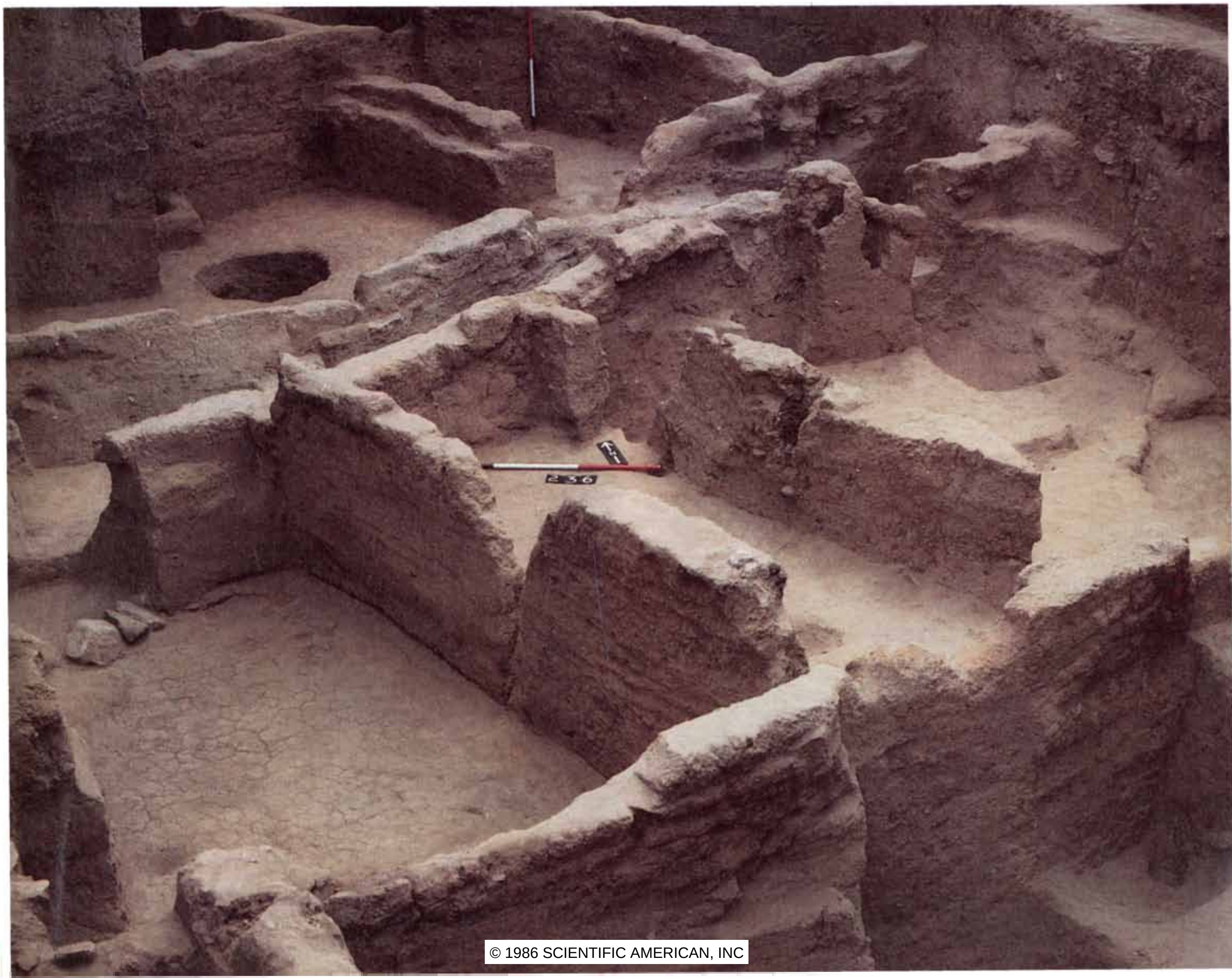
Although the settlement was reconstructed periodically, the methods of building employed by its residents seem to have remained more or less the same through much of the Bronze Age. As a result the village on the mound may have changed little in ap-

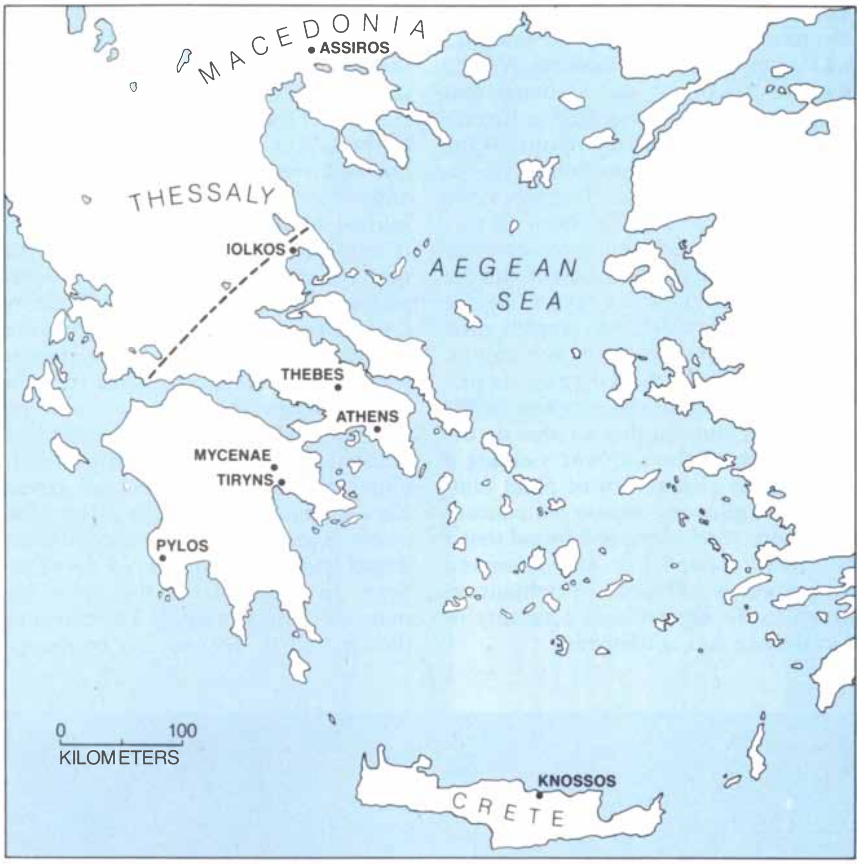
BRONZE AGE WALLS were uncovered by the authors in a large mound near the modern Greek village of Assiros. They have timber frames that are filled in with bricks made of sundried mud. The walls formed a honeycomb of living quarters and storage rooms separated by yards and alleyways; the entire settlement was enclosed by a defensive circuit wall made of mud bricks. The rooms shown in the photograph were occupied by about 1200 B.C., in the late Bronze Age. Just under them are the floors of crop storerooms that were destroyed by fire in about 1350 B.C. From the charred rubble on the floors has come a rich haul of botanical remains consisting of burned fragments of stored crops. The circular pit at the rear was for storage; the depressions at the right are the result of later digging in the Iron Age.

pearance with each successive rebuilding. The circuit wall was constructed of clay and of bricks made from sun-dried mud. Within that wall stood rectangular rooms, which had timber frames filled in with mud brick. (The timber frames may have been intended to strengthen the rooms against the earthquakes that plagued Macedonia in the Bronze Age and continue to do so.) Interspersed among the rooms, which served as dwellings and for other purposes, were yards. Groups of rooms and yards were divided by alleys, giving the settlement a rectilinear plan. The alleys were surfaced with gravel but were too narrow to allow even a pack animal to be driven along them, much less a cart, and so it must be assumed they were mainly for walking.

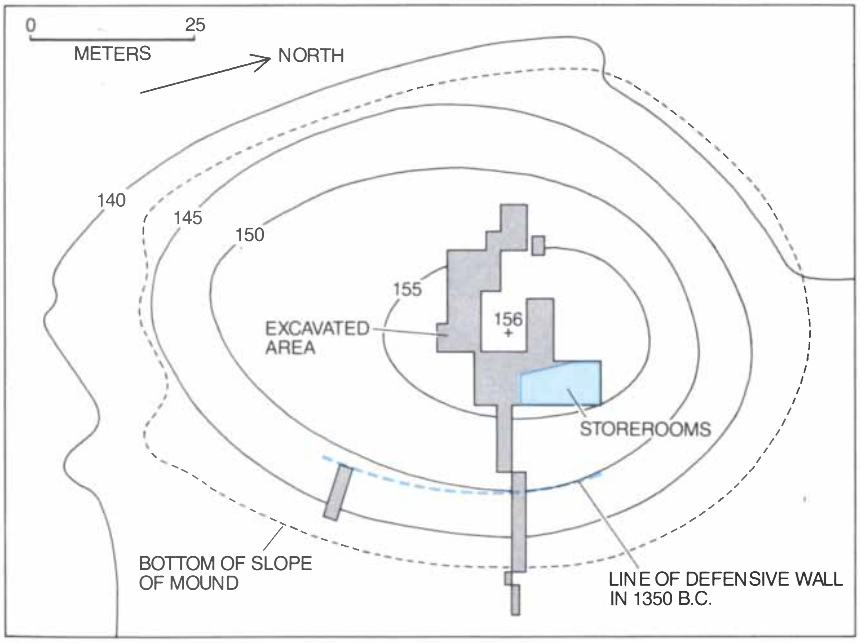
In the northern part of this honeycomb of rooms, yards and walkways we uncovered the storerooms. The discovery, which is in some ways the most stimulating of our findings, was completely unforeseen. At the start of the dig it was assumed that Assiros Toumba concealed a Bronze Age site, but neither the nature of the settlement nor the duration of its occupation was known. Therefore the original plan was to dig about 10 percent of the mound to a depth of three meters in order to document the sequence of occupation layers. In the course of that work we sampled each layer systematically for plant remains. In 1977, after the third digging season, study of the botanical remains in the laboratory showed that an area on the north edge of the site was yielding a particularly rich return of plant matter. The following season we concentrated our work there and found that a fortunate coincidence had preserved an intriguing collection of remains relating to the agricultural economy of the Bronze Age settlement.

It turned out that in about 1350 B.C. a fire had destroyed several rooms of the settlement. Scattered on the floors of the burned rooms were burned mud bricks and the charred remains of roof timbers. The people of Assiros had not bothered to clear the wreckage and reconstruct the floors of the burned buildings. Instead they merely smoothed the debris and used the newly leveled surface as the underpinnings of a new floor. The framing timbers from the burned building were cut near the base, and the stumps were used as foundations for the new rooms. As a result of this "new on old" method of building, a substantial amount of grain remained in the debris above the old floor. Charred grains were scattered through the fallen mud brick. Even more grain was concentrated in a layer on the old floor; indeed, in some places the grain lay many centimeters deep. Three rooms (designated 9, 12 and 14) contained





ASSIROS AND MYCENAEAN SOCIETY had considerable contact during the period of the Bronze Age palaces. Assiros is in Macedonia, which now forms part of northern Greece. At the height of the age of palaces (from 1400 to 1200 B.C.) Macedonia constituted a separate cultural region; the northern border of Mycenaean society (*broken line*) lay south of Macedonia in Thessaly. There were Mycenaean palaces at Mycenae, Tiryns, Pylos and Thebes and perhaps also at Iolkos and Athens, although the evidence there is less clear.

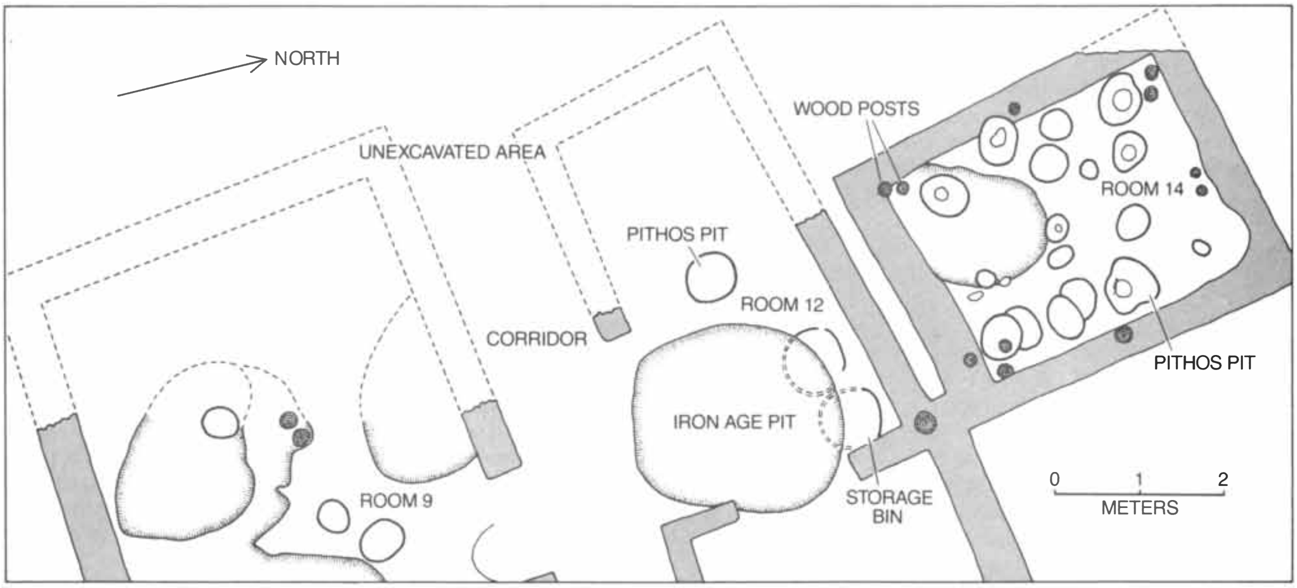


ASSIROS TOUMBA is the mound where the remains of the Bronze Age settlement were found. The mound, some 14 meters high, resulted from the building and rebuilding of the settlement 20 times or more between 1800 B.C. (in the middle Bronze Age) and 800 B.C. (in the early Iron Age). The excavated area is shown in gray, the crop storerooms in color.

thick layers of charred debris, as did an adjoining covered passageway. A fourth room and an adjacent yard seem to have been part of the same architectural unit, but they were not destroyed in the fire and therefore did not yield the same rich haul of botanical remains.

The botanical and nonbotanical material found in the complex left little doubt that it comprised a set of crop storerooms. Among the nonbotanical remains were traces of a variety of objects associated with agricultural storage. In the floor of Room 9 at least six pits, and perhaps more, had been cut. The pits, which are approximately 30 centimeters deep, nicely match the pointed bases of clay jars found elsewhere on the mound. Such jars, which were common Bronze Age storage vessels, are known in Greek as *pithoi* (the singular is *pithos*). The floor of Room 14 was better preserved, and it had more than 10 *pithos* pits. Although a piece of a clay disk that might have been the lid of a *pithos* was discovered in one room, few fragments of the jars themselves were found. It seems likely that the storage jars were not harmed in the fire and were salvaged. Much of the floor of Room 12 had been disturbed by digging in the Iron Age, and no indentations for storage jars were found there. That room, however, did contain two storage bins whose willow frames were covered with clay.

After considering the evidence for the number and type of containers, we were able to estimate the storage capacity of the complex, or at least of the three rooms where burned debris was found. It appears that rooms 9, 12 and 14 were employed exclusively for storage and that the crops were put in *pithoi*, clay bins and perhaps also sacks or other perishable containers that have not sùrvived. In making our computation we assumed that the placement of the *pithoi* on the floor of each storeroom followed the spatial pattern observed on the intact part of the floor of Room 9, where six pits were found. On the basis of these assumptions, and taking into account the fact that each *pithos* could hold between 100 and 150 liters, it was calculated that rooms 12 and 14 had a storage capacity of about 1,000 liters each and Room 9 a capacity of from 2,000 to 3,000 liters. The total volume of between 4,000 and 5,000 liters is more than enough to feed 20 people for a year. The concentration of that much storage capacity in one area of the mound is remarkable for prehistoric Greece, where the dispersal of storage facilities in individual dwellings is the rule. The only parallel for the concentrated arrangement seen



STORAGE COMPLEX included three rooms that were destroyed by fire. In the floors were pits that could each have held the base of a storage jar called a *pithos*. Room 12 also contained the remains of two storage bins that had willow frames covered with clay. After the fire the residents salvaged the jars, leaving the grain. The burned debris was then smoothed to form the base of a new floor. In some places the grain lay many centimeters thick on the old floor, providing the authors with excellent samples of the Bronze Age crops.

at Assiros is to be found in the Bronze Age palaces of southern Greece.

In the past an excavator who found storerooms like these would probably have assumed that the grain scattered about had been so thoroughly disturbed during and after the fire that it could not possibly yield any information about the original pattern of storage. At Assiros, however, detailed sampling enabled us to infer quite a lot about the number of stored crops and their spatial arrangement. Most of that information came from the old floor, which lay under the layer of charred debris. Although the grain mixed with the debris had been severely disturbed, the grain lying on the floor apparently had not. We found many areas where the scattered grain included a very high proportion of seeds from a single plant species. Those concentrations of one species, we reasoned, must have lain undisturbed after the fire. Any serious disturbance would undoubtedly have mixed the contents of several containers, resulting in random mixtures of species rather than concentrations of one type. Therefore we inferred that these concentrations held significant information about the storage pattern. Each one was analyzed individually. Samples were taken from the center and the periphery of the concentration as well as from the area between two concentrations. If a concentration was large, it was sampled many times.

The botanical sampling first yielded the information that the residents of Assiros had cultivated at least seven species of plants. Three species of wheat were conclusively identified: einkorn (*Triticum monococcum*), emmer (*T. dicoccum*) and spelt (*T. spelta*). The seeds of two other species—bread wheat (*T. aestivum*) and macaroni wheat (*T. durum*)—are indistinguishable. Since grains of this type are present, it is possible that both species were grown; on the other hand, perhaps only one of the two species was cultivated. At least two cereals were grown in addition to wheat: six-row hulled barley (*Hordeum vulgare*) and broomcorn millet (*Panicum miliaceum*). There was also a pulse called bitter vetch (*Vicia ervilia*). (The pulses make up the group that includes peas and lentils.) The seven species probably supplied most of the diet of the Bronze Age inhabitants. The cereals may have been consumed as bread or as gruel.

The potential of spatially intensive botanical sampling such as that done at Assiros extends far beyond providing a catalogue of plant species. Among the other things it can show is which crops were grown and stored individually. It is clear from the almost pure samples of einkorn, millet and bitter vetch that these three species were stored—and therefore presumably grown—as separate crops. The overwhelming predominance of bread (or macaroni) wheat or hulled barley in some samples suggests these species were also grown separately. The case of emmer and spelt is more complex. Although the ratio of the two species varies from sample to sample, nowhere is one found without the other. It appears emmer and spelt were grown and stored together. In this they resemble the "maslin," or mixed, crops of wheat and barley grown today by some Greek peasant farmers. Since the two plants thrive in slightly different conditions, combining them reduces the chance of complete crop failure.

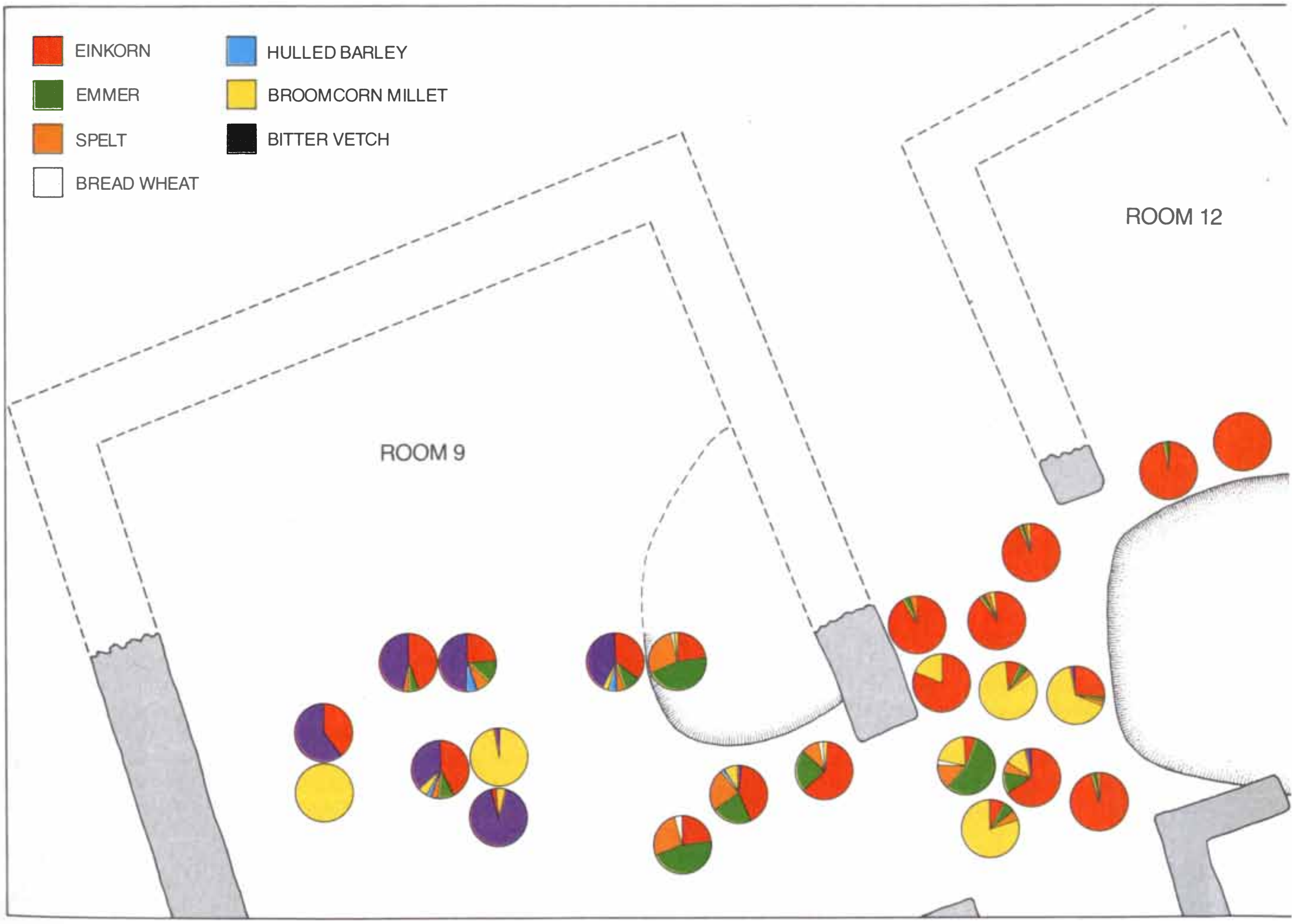
The overall diversity of crops grown at Assiros probably also had a beneficial effect. Each of the six crops for which evidence has been found in the storerooms (einkorn, millet, bitter vetch, bread—or macaroni—wheat and barley in addition to the mixture of emmer and spelt) is sown and harvested on a different schedule. Therefore cultivating such a range of crops spreads the periods of intense agricultural labor throughout the year. In addition each crop has slightly different requirements in terms of climate and rainfall, and growing a variety of them can give protection from the vicissitudes of temperature and rainfall. The latter may have been particularly significant for the residents of Assiros, because the eastern part of Greece has an arid climate in which rainfall can be quite capricious.

The six crops not only were cultivated on different growing cycles but also were probably stored in different forms. Gordon C. Hillman of the University of London has studied how crops are processed today in simple agricultural settings, including Turkish villages. Because such economies are similar to those of Bronze Age Greece, Hillman's work can help to show how crops were processed in prehistoric times. His findings suggest that some plants are generally fully processed before the useful portion is stored. At ancient Assiros that group would have included bitter vetch, barley, millet, bread wheat and macaroni wheat. The processing of those plants requires only a single threshing to detach the seeds from the rest of the plant, followed by winnowing and sieving to retrieve the seeds. What goes into the storage bins is the grain, ready for cooking.

Another group of plants requires two phases of processing, and the crop can be stored at more than one stage. In that group are einkorn, emmer and spelt. The seeds of those plants are tightly enveloped by thick husks called glumes, and the three species are known as glume wheats. Together the grain and its covering are called the spikelet; the spike, or ear of the grain, is made up of many spikelets attached to the central stalk. For the crop to be transformed into food the wheat must first be threshed to break the ear into individual spikelets. The mixture is then winnowed to retrieve the spikelets, leaving behind the inedible remainder of the plant. The spikelets are threshed a second time to separate the wheat grains from the chaff, which consists mainly of the husks. The grain is then obtained either by a second winnowing or by sieving.

The two-stage process affords two different points for storage. The crop can be stored after the first threshing as intact spikelets. Alternatively, the second threshing can be carried out and the crop stored as fully processed grain. Storing the spikelets is advantageous on two grounds. Threshing is arduous. Putting the spikelets in storage and then threshing them as the grain is needed spreads the work of threshing through the year. Moreover, the spikelets are more resistant to the attacks of insects and fungi than the grain is. The residents of Assiros apparently exploited these advantages and stored some of their glume wheat as spikelets; we found glume-wheat samples containing intact spikelets and also samples with very high proportions of chaff.

So far we have described how spatially intensive sampling of plant remains has yielded information about the number and type of crops cultivated at Assiros and also about how the farmers there processed and stored what they grew. Can similar techniques be used to answer more specific questions, such as how many jars of each product there were in a particular room of the complex? The answer is that they can. At least they can provide an estimate of the minimum number of storage containers of each crop in a particular area of the complex. For example, if in a corner of one room two concentrations of barley are found to be separated by a concentration of einkorn, it follows that there must have



SAMPLES OF GRAIN from the floors of the storerooms yielded considerable information about crop storage at Assiros. The seeds of at least seven species have been identified. Among them are four types of wheat: einkorn, emmer, spelt and bread wheat. There were two other cereals—hulled barley and broomcorn millet—and bitter vetch, which is a relative of peas and lentils. Each circle indicates

been at least three containers in that corner: two filled with barley and one filled with einkorn.

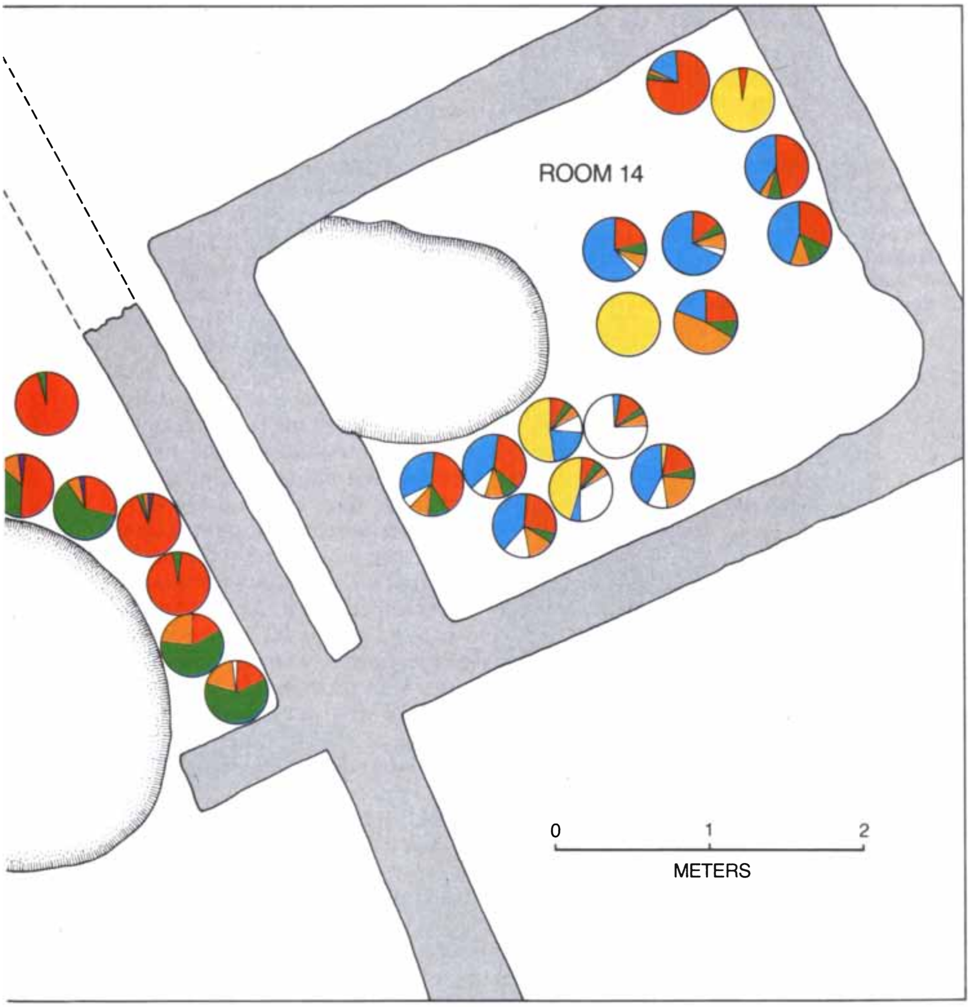
The reason such an analysis yields only the minimum number of storage vessels is that the concentration of, say, einkorn may contain a mixture of grains from two jars of einkorn that were adjacent when the fire broke out. By combining the minimum estimate with a count of the *pithos* pits, however, we can approximate a complete count of the containers of each grain type. The clearest instance of this method applies to Room 9, where seven separate concentrations were found on the preserved part of the floor: two concentrations of einkorn, two of emmer with spelt, two of broomcorn millet and one of bitter vetch. As we mentioned above, at least six pits for *pithoi* were found in the floor of Room 9. It seems reasonable to conclude that there were at least seven storage containers in the preserved area of Room 9 and that the list of concentrations reveals the contents of those seven or so jars. Similarly, the floor of Room 14 revealed eight concentrations of grain and perhaps as many as 10 *pithos* pits. It is more difficult to be certain of the distribution of containers here than in Room 9, since one or more of the concentrations of grain may have come from several containers. Nevertheless, the sampling procedure, when combined with an examination of the nonbotanical remains, can give a plausible description of the storage pattern in each room.

Analysis of the storage pattern in the individual rooms reveals the striking fact that each of the three storerooms held a different range of crops. The list of concentrations given in the preceding paragraph for Room 9 excludes bread (or macaroni) wheat, since grains of that type were not found there. Indeed, bread (or macaroni) wheat was stored only in Room 14. Bitter vetch, on the other hand, was put only in Room 9. Millet is absent from Room 12. The uneven distribution of products is quite significant for understanding how the storerooms functioned. Several factors must be taken into account, however, before the full significance of that pattern becomes apparent. The first step is to note that the capacity of the storage complex is much greater than the capacity required for a single household. Since that is so, one must conclude that either each storeroom belonged to a different family or the complex constituted a communal storage area.

There is good reason to think the storerooms were not used by individual families. The list of crops in each storeroom omits at least one plant with highly desirable properties. For example, barley, which is commonly grown today in the Greek islands as a low-risk crop, is absent from rooms 9 and 14. Millet (absent from Room 12) is sown in what is otherwise a slack period for farm work. Vetch (absent from rooms 12 and 14) is quite useful in schemes of crop rotation. Therefore any family that grew the range of crops represented in any one storeroom would have added considerably to its labor by failing to take advantage of the complementary properties of the full range of crops. Furthermore, the families that used rooms 12 and 14 would have had a diet that was far less than optimum. Vetch, which is absent from those rooms, is an excellent nutritional complement to cereals.

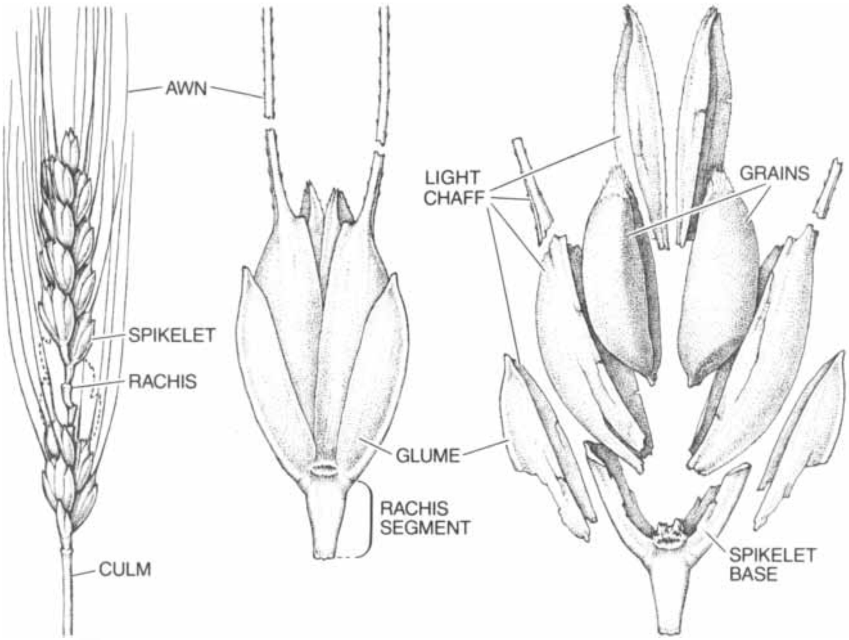
Suppose, on the other hand, that the three storerooms formed part of a single unit. If that were so, the group controlling the complex would have been exploiting the complete range of available crops. Support for the idea that the rooms formed a unit comes from the architectural record. The passages connecting the storerooms, unlike most of the alleys in the settlement, were not covered with gravel. The absence of covering suggests the passages were originally under roofs that joined the rooms in one edifice. Yet who could have controlled that structure? The size and capacity of the storerooms suggest only the community as a whole, or a substantial part of it, could have been in possession of the crops stored there. It appears what we have found at Assiros is a truly collective phenomenon, which must have required a good deal of cooperation among the inhabitants.

It should not be supposed that the products stored in rooms 9, 12 and 14 necessarily exhaust the storage facilities at Assiros. Individual families probably stored food in their own dwellings. Rather than providing the

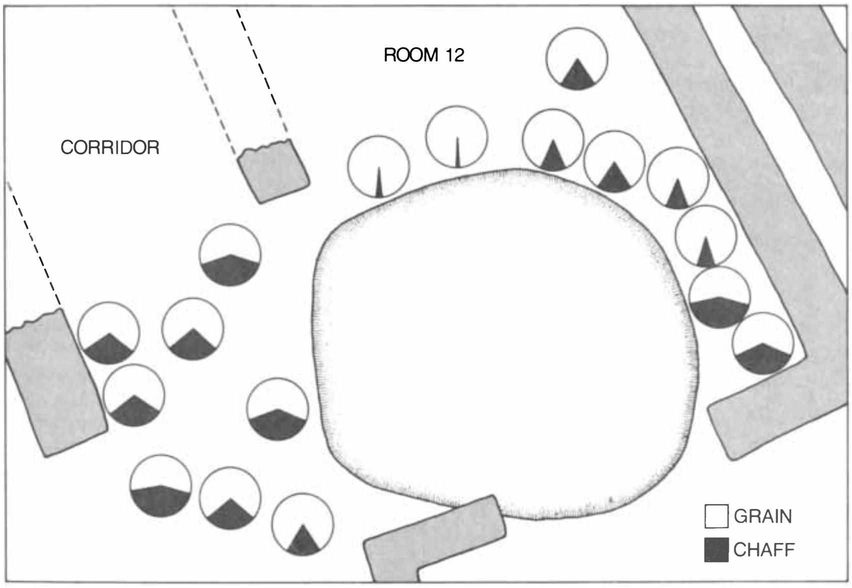


the site of a sample and gives the proportions of the seven species in that sample. The fact that some samples are heavily dominated by one species suggests the grain has not been much disturbed since 1350 B.C. Consequently sampling can reveal the pattern of storage.

total storage capacity of the community, the complex we found was probably an area where a reserve of food was kept as a kind of communal insurance policy against poor harvests or complete crop failure. Such a surplus may have been collected as obligations in kind from the community, and the stored produce may have been returned to individual community members for services in rebuilding the circuit wall or other communal tasks. The exact nature of the social mechanisms accompanying such collective storage cannot yet be reconstructed. Yet the impression of cooperation is a strong one, and it helps to create a picture of what the settlement was like in the Bronze Age.



GLUME WHEATS are so called for the glumes, or husks, that surround the grains. The crops grown at Assiros included three species of glume wheat: einkorn, emmer and spelt. Emmer is shown in the illustration. At the left is the spike, or intact ear. Each spikelet, containing a pair of grains, is attached to the rachis, the part of the central stalk that lies within the ear. In processing glume wheats the crop is threshed once to break the spike into spikelets (*middle*). A second threshing separates grains from chaff (*right*). The chaff consists of light fragments and the tough bottom parts of the glumes, which form the spikelet base.



RATIO OF CHAFF TO GRAIN in glume-wheat samples yields information about how those crops were stored. Glume wheat can be stored as intact spikelets after one threshing or as fully processed grain after two threshings. If it is stored after only one threshing, most of the chaff still surrounds the grain. The high proportion of chaff in samples from Room 12 shows that at least some of the glume wheat at Assiros was stored after a single threshing.

The capacity for coordinated activity would seem to set Assiros above the level of a simple agricultural village. Beyond that observation, however, it is not easy to specify the economic and social structure of the community on the mound. One promising model for understanding the settlement in its relation to the surrounding population is that of the redistributive system. Redistributive systems are known to have existed in early Near Eastern societies such as Sumer and Ugarit during the Bronze Age; the network of relations around the Minoan and Mycenaean palaces has also been interpreted as a redistributive system.

In a redistributive system the authorities at the center, which may be a palace, a temple or a town, have the right to expect certain goods and services from the local populace. In return the authorities are expected to provide material benefits, such as food or raw materials, and less tangible benefits, such as military protection or the observance of ritual. A crucial feature of such systems of reciprocal obligation, which have been documented in clay tablets found in Near Eastern temples and at Knossos and Pylos, is that they do not include money or markets. The expectations of each side are governed by customs that make it possible for goods and services to flow freely without monetary exchange or barter.

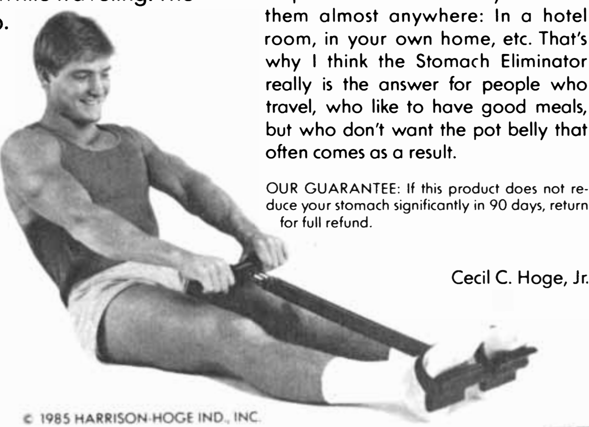
Assiros Toumba was clearly a much simpler place than the great Bronze Age palaces. Perhaps it is best understood as "protopalatial": the kind of settlement that might, in the right circumstances, turn into a fully developed redistributive center. If that is the case, one can make certain predictions about what we shall find in the course of our further work. Any redistributive center, no matter how small or primitive, must have a network of connections with surrounding settlements. To discover what settlements lay within the orbit of Assiros Toumba we plan to survey the surrounding area on foot.

Furthermore, the center generally holds only a fraction of the population of a redistributive system. If Assiros Toumba was such a system, perhaps only a small proportion of the population needed to build and rebuild the circuit wall lived on the mound itself. The rest of the population would have lived in the outlying settlements and contributed their labor to the raising of the defensive wall in exchange for protection or food or other benefits. It is possible that the population of Assiros Toumba itself numbered only between 50 and 100 individuals. Further digging within the circuit wall will be needed to find out whether this was the case.

Redistributive systems are currently the focus of a lively theoretical interest among archaeologists and anthropologists, and the work at Assiros Toumba could ultimately make a useful contribution to that discourse. One significant question is how redistributive systems originated. Although the question has not been answered, an intriguing hypothesis is that they protected agricultural communities against famine by providing for the storage of a surplus. Some impressionistic evidence for that hypothesis is provided by the biblical story of Joseph. In heeding his dream and storing the surplus of the seven fat years to provide for the seven lean years, Joseph was fulfilling the obligations of the center in a redistributive system.

If we find the storeroom complex supplemented rather than replaced storage in individual households, it would support the notion that redistribution functioned to damp the oscillations of the agricultural economy. The reason is that it would then seem the storage complex was intended to hold a reserve and not simply to be a centralized system of collective storage for everyday use. In this way the work at Assiros Toumba might illuminate some important theoretical questions. That possibility, however, lies in the future.

For the present it should be said that the work on the mound has already far exceeded our expectations. We began our excavation knowing little about what remains there were at Assiros. It seemed reasonable to think we would find the remains of a typical Bronze Age village containing some 40 to 50 households. We expected those households to be independent entities, relying on their own labor and lacking the social mechanisms needed to bind them into a larger community. Instead we found a community that apparently had complex mechanisms for social cooperation, including perhaps a redistributive system that subsumed both labor and agricultural surplus. In this light Assiros becomes far more interesting, because it resembles the great palace towns of southern Greece as much as it does a simple agricultural village. In the years to come much more will be learned about how the palace economies of the Bronze Age originated, and Assiros will undoubtedly make a full contribution to that enterprise.



Athletic Clothing

The thorough attention now given to the design of athletic equipment has contributed to new records in the speed sports and to enhanced protection or performance in other contests

by Chester R. Kyle

A 50-year-old photograph of an athlete brings sharply to the viewer's eye the change that has taken place in that sport's clothing. A football player wears a skimpy helmet, loose-fitting pants and light padding, a baseball player has on a baggy wool suit and a woman plays tennis in a long skirt. It is evident that little consideration was given to the suitability of the clothing for the sport. Today a good deal of thought goes into the design of clothing for athletes. Indeed, in the speed sports a properly designed outfit can provide the winning margin in a close race. In other sports the right clothing can make a distinct contribution to performance or increase the athlete's protection from injury.

Each sport has distinctive requirements for clothing and equipment. It would be impossible to explore them all here and to describe the improvements that have been made by utilizing new materials and approaching design from an engineering point of view. Instead I shall focus on a few examples: aerodynamic clothing and equipment, the running shoe and the helmet.

Aerodynamic clothing and equipment are of great importance to skiers, speed skaters and cyclists and to athletes who compete in luge and bobsled races. Speeds are high enough in these sports to make wind resistance the major retarding force. Tests in wind tunnels and in the field have shown that a reduction in aerodynamic drag will improve performance.

Aerodynamic drag on the human body can be lowered in three ways. The commonest way is to change the position of the body in relation to the wind. The uncomfortable crouched posture of downhill skiers, speed skaters and cyclists reduces the area of the body facing the wind and makes a more streamlined form. A diver exhibits the ultimate in streamlined posture and low frontal area. This extreme position is not practical in most sports, although the nearly prone luge sledder approaches it.

The second way of dealing with drag is to design equipment that helps to streamline the body. An example was the helmet worn by members of the U.S. cycling team in the 1984 Olympic Games; it changed the shape of the upper part of the head to resemble the canopy on a jet aircraft.

The third method is to make clothing as smooth and tight as possible to minimize friction arising from contact with the air. That is why tight, smooth, aerodynamic costumes have become standard in the speed sports. The effect is evident in speed skating and cycling because an overall reduction of from 6 to 10 percent in wind drag is possible if the competitor wears aerodynamic rather than conventional clothing.

Beginning in 1982 Paul VanValkenburgh, Joyce S. Kyle (my wife) and my colleagues and I designed and tested clothing, helmets and bicycle components for the U.S. cycling team. We found that at 30 miles per hour on smooth, level pavement wind resistance accounts for about 90 percent of the total drag force on a bicycle. Rolling resistance of the tires and friction on the bearings account for the remaining 10 percent. About two-thirds of the aerodynamic drag force is due to the rider. Thus the greatest chance for gains in speed lies in improving the aerodynamics of the rider. It was this finding that led VanValkenburgh to design the bicycle helmet.

In tests in the low-speed wind tunnel at Texas A&M University we found that our prototype clothing and helmet, compared with the best cycling outfits then available, could lower the overall wind drag by as much as 6 percent. In the 4,000-meter pursuit race the reduction in drag would mean an improvement of as much as three seconds in the record time. Indeed, American cyclists wearing similar clothing have won many medals in international competition.

VanValkenburgh, Peter R. Cavanagh, Jack Lambie and I also worked on streamlined cycling shoes. We calculated that they would take another 1.5 seconds off the 4,000-meter record. Up to now they have not been worn in competition because of difficulties in adapting the shoe to the human foot and the possibility that the shoes may be illegal according to international rules that prohibit most forms of streamlining in standard-bicycle races.

In track and field it had not been shown until recently that clothing affects running speed, even though many workers had pointed out the importance of wind resistance in running. In middle- and long-distance running about 6 percent of the total energy is expended to overcome wind resistance. In Olympic events the maximum running speed ranges from about 12 m.p.h. in the marathon to 27 m.p.h. in the 100-meter dash. Recent wind-tunnel tests I carried out showed that reductions of from 2 to 6 percent in aerodynamic drag are easily achieved by simple changes in the shape a runner presents to the wind.

Loose jerseys and thick or long hair appear to be the primary sources of drag. Even long, rough socks generate 1 percent more drag than bare legs. A tight cap will reduce the wind resistance by about 4 percent. Swapping loose, rough clothes for tight, smooth ones will gain a further reduction of 4 percent.

That reductions in wind resistance will improve running times is hard to verify: the predicted differences in time are small and fall within the area of the variability a runner can exhibit from race to race. In one case, however, the effect of reduced wind resistance has been demonstrated. In Mexico City, where the altitude is 2,255 meters (7,400 feet), the density of the air is about 20 percent lower than it is at sea level. The wind resistance is therefore about 20 percent lower too. A. J. Ward-Smith of Brunel University in England has shown that sprint speeds are an average of about 1.7 percent higher in Mexico City than they are at sea level.

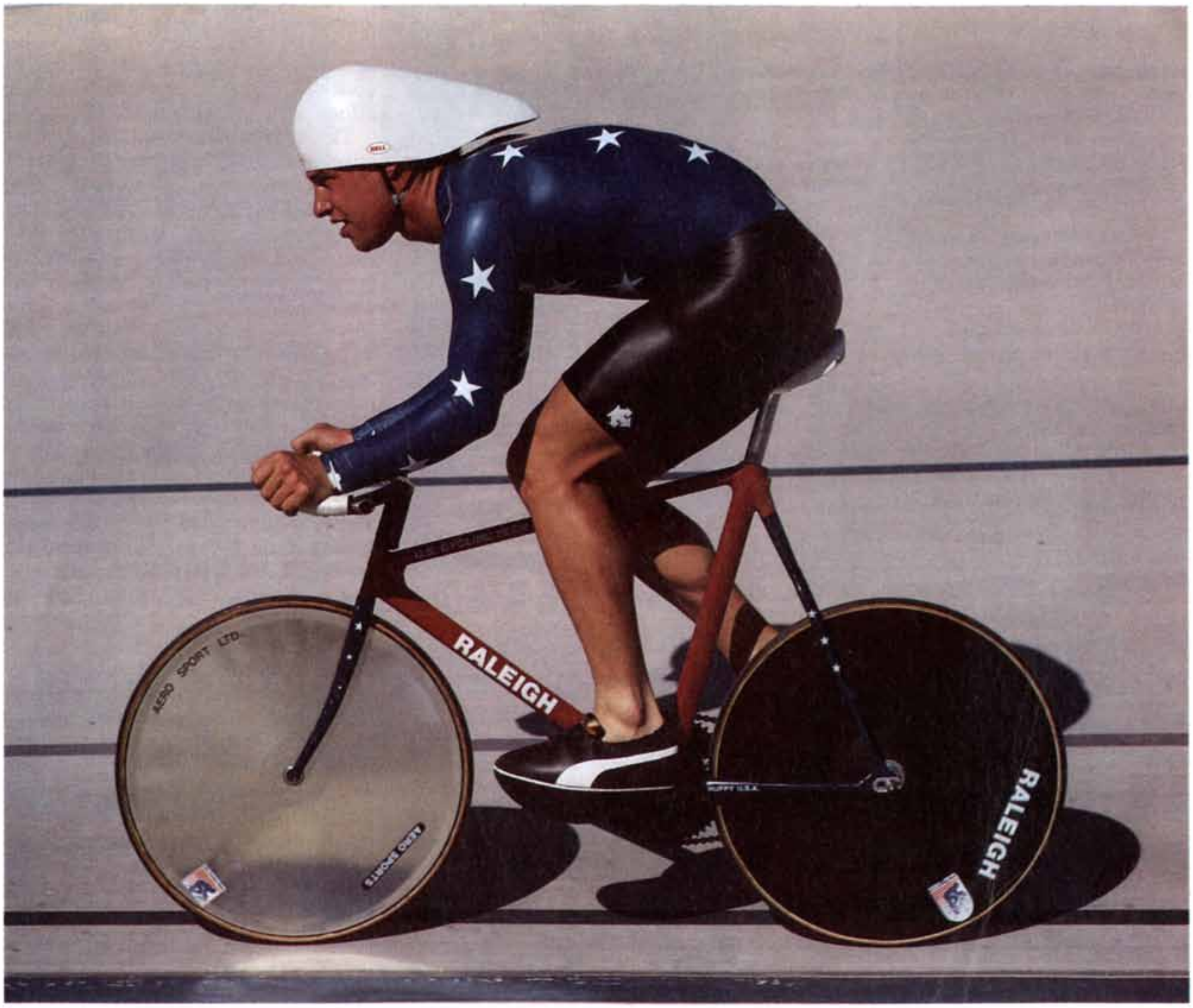
Employing mathematical models of sprinting and distance running, I have calculated that a drop of 2 percent in wind resistance can significantly change the length of a winner's lead. In a race between runners of equal ability the winner's lead varies from about four inches in the 100-meter dash to more than 30 yards in the marathon. On the international level winning margins are often less because the athletes are very closely matched. An improvement in clothing can easily give a competitive advantage. Tight running costumes similar to those in the other speed sports were introduced recently and are being widely adopted.

Special footwear for running is also a fairly recent development compared with the antiquity of the sport. Formal track and field meets began at Eton in 1837, and the first recorded dual meet was between Oxford and Cambridge universities in 1864. A pair of spiked running shoes from that era is preserved in the Central Museum and Art Gallery in Northampton, England. By 1894 the spiked shoe made of light leather had assumed a form that remained almost unchanged for more than 60 years.

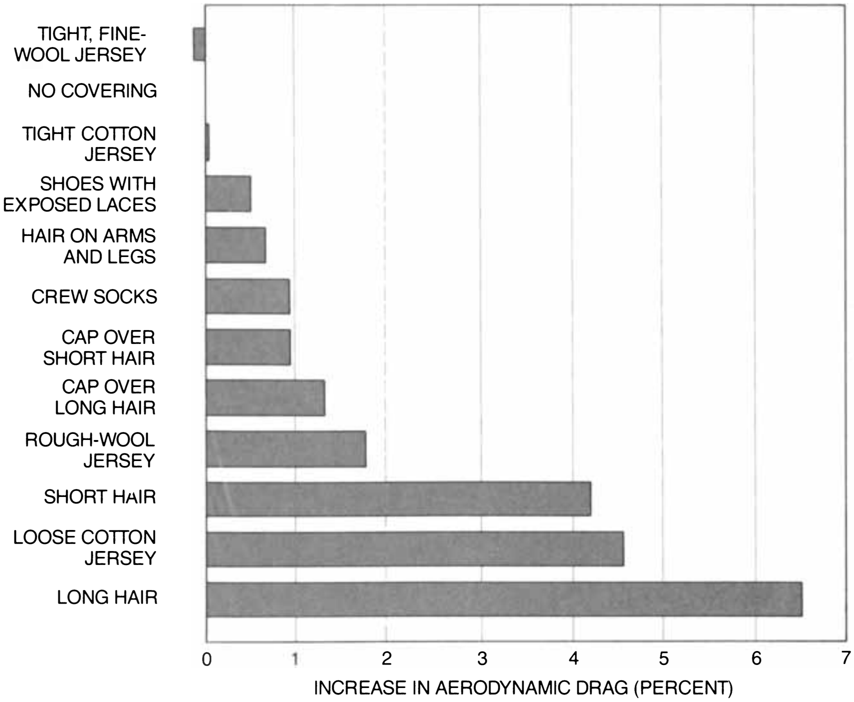
The first shoes for long-distance running had high tops of leather and heels and soles of leather or rubber; they did not differ much from standard walking shoes. Shoes for both distance running and sprints began to change significantly in the late 1950's, but it was not until the 1970's that competition and new technology brought about a rapid proliferation of new designs.

In large part the radical changes in running shoes result from what has been learned in the past decade about the basic mechanics of running. Experts in biomechanics have employed high-speed photography, video cameras, pressure plates, load cells, devices for measuring oxygen consumption and other instruments to study in detail the important variables in running. The variables include the motions of the feet and legs, the time for each segment of motion, the forces on the runner and the level of energy required by various types of running.

This work has revealed that several critical changes take place as running



AERODYNAMIC EQUIPMENT designed by the author and his colleagues for bicycle racers includes streamlined shoes and a helmet shaped like the canopy of a jet aircraft. They and the tight clothes reduce the aerodynamic drag on the rider by as much as 10 percent. The rider, photographed in the Los Angeles Velodrome, is Steve Hegg, who won a gold medal in cycling at the 1984 Olympics.



AERODYNAMIC EFFECT of various kinds of clothing on a runner is charted. The wool jersey ranks ahead of no covering because it functions like the dimples on a golf ball.

speed goes up. They include the action of the legs and feet, the forces on them and the metabolic costs of running.

As speed increases, the frequency and the length of the stride increase. So does the flight time: the periods when both feet of a runner are off the ground. A sprinter running flat-out can have both feet off the ground more than half of the time.

As speed increases, the center of pressure of the foot's contact moves forward. Usually someone running slowly begins a foot strike on the rear outside edge of the foot, whereas a sprinter lands on the forward outside edge. Yet about 20 percent of distance runners are forefoot strikers. Cavanagh, one of the pioneers in the detailed study of the running shoe, has ascertained in his work at Pennsylvania State University that the American marathoner Bill Rodgers does not touch the ground with the rear of his foot (when the course is level). He could wear a heelless shoe.

The forces entailed in running become more significant with increasing speed. The velocity at which the foot strikes the ground increases, as does the maximum force of the foot's contact with the ground. The energy lost in the shock of contact also increases. The angle and the rate of flexion of the knees and hips change with higher speeds. This is probably an automatic adjustment made to help absorb the heavier shock without injury.

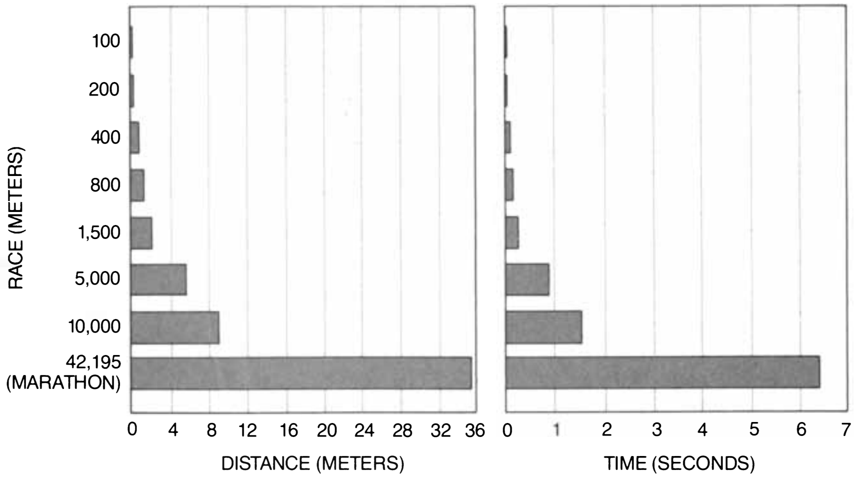
At higher speeds the foot supinates, or rolls outward, increasingly before striking the ground. After striking the ground the foot pronates, or rolls inward, and the center of pressure moves forward and toward the midline of the foot. Because of the high pressure and force generated, shock absorption becomes one of the most important functions of the running shoe. Workers at Nike, Inc., have found that the maximum force on the foot after ground contact is as much as three times the weight of the body and that the acceleration transmitted to the leg can be 10 times greater than normal gravity.

The metabolic cost of running rises linearly with speed. As a result, since the average speed decreases as the distance of the run increases, the energy consumed remains between 70 and 90 kilocalories per kilometer.

These findings and additional ones have given rise to a variety of track and running shoes. A sprinting shoe is extremely light and has fine spikes. They are necessary because the traction force needed by a sprinter approaches the weight of his body. (In distance running the typical traction force is about 40 percent of the body weight, so that spikes are not necessary.) The spikes are designed to minimize energy losses while damaging the running surface as little as possible. Since a sprinter lands on the outside of the foot, traction ridges are placed on the side of the shoe to prevent slipping. The high-jump shoe has spikes on the takeoff foot; they are positioned according to the athlete's jumping style. The opposite shoe is much lighter and has no spikes.

The modern shoe for distance running has several features to reduce injury and increase performance. It has a midsole made of expanded foam to absorb shocks. The raised heel wedge is designed to reduce stress on the Achilles tendon, thereby preventing tendinitis there. The flare of the sole, the stiff heel counter and the variable hardness of the sole material all help to control excess motion of the foot after it touches the ground.

Some of these components are still fairly soft. Soft materials allow the foot to shift and to pronate excessively. This kind of movement is thought to be the cause of some knee injuries. Consequently much of the current ef-



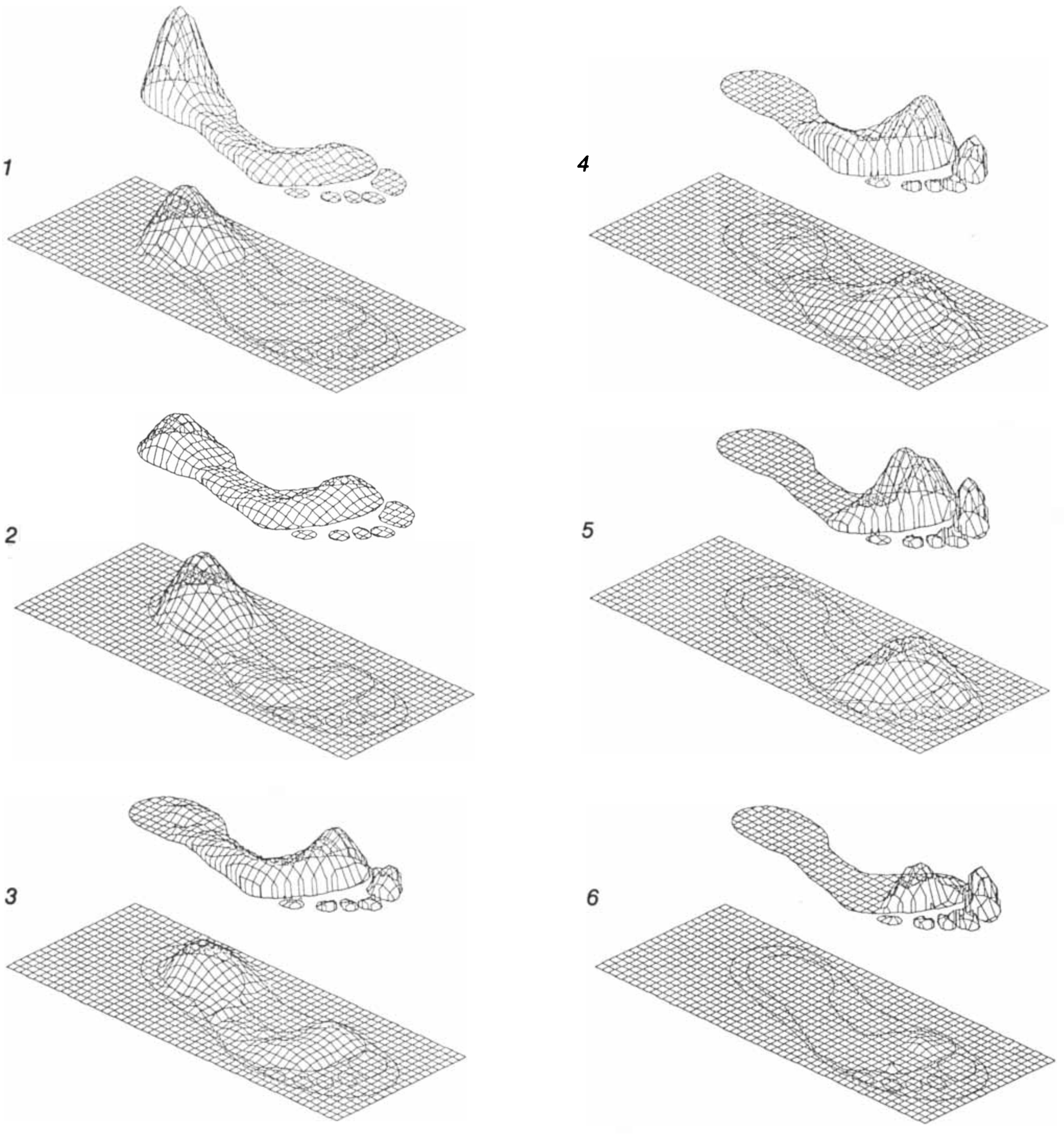
DECLINE IN WIND RESISTANCE can increase the lead of the winning runner in a race and decrease the winning time. The table reflects the results of a 2 percent reduction in the resistance arising from wearing aerodynamic clothing or from running at a high altitude.

fort in the design of running shoes is aimed at controlling excessive pronation. The hardness of the outsole and the pattern of the tread can be designed to keep the foot from rolling inward. Moreover, the inner part of the heel wedge can be made harder than the outer part. This feature and the arch-supporting inner liner serve together to control excessive pronation.

The bottom of the shoe is made of a harder and more durable material than the midsole. It can be specifically designed for traction, flexibility, long life, foot control, shock absorption and several other requirements. Inside the shoe a soft, flexible inner liner is shaped to give support to the arch —a feature that was once incorporated only in orthopedic running shoes.

The upper part of a running shoe serves no purpose except to hold the sole on the foot. If it were possible to attach the sole to the foot with a nonirritating adhesive, the upper could be eliminated. Such a bizarre shoe, however, would probably last for only one race and might cause injuries.

The upper, which usually accounts for about 30 percent of the typical run-



RUNNING TEST shows the pressure generated by a runner's bare right foot on a hard surface and on a foam pad during a period of about 500 milliseconds after the foot makes contact with the ground. The height of the grid lines is proportional to the contact pressure. In each case the reading from a hard surface appears above the reading from the pad; the times are comparable. The pad, like a running shoe, spreads the pressure and absorbs shock. The tests were made by Peter R. Cavanagh of Pennsylvania State University.

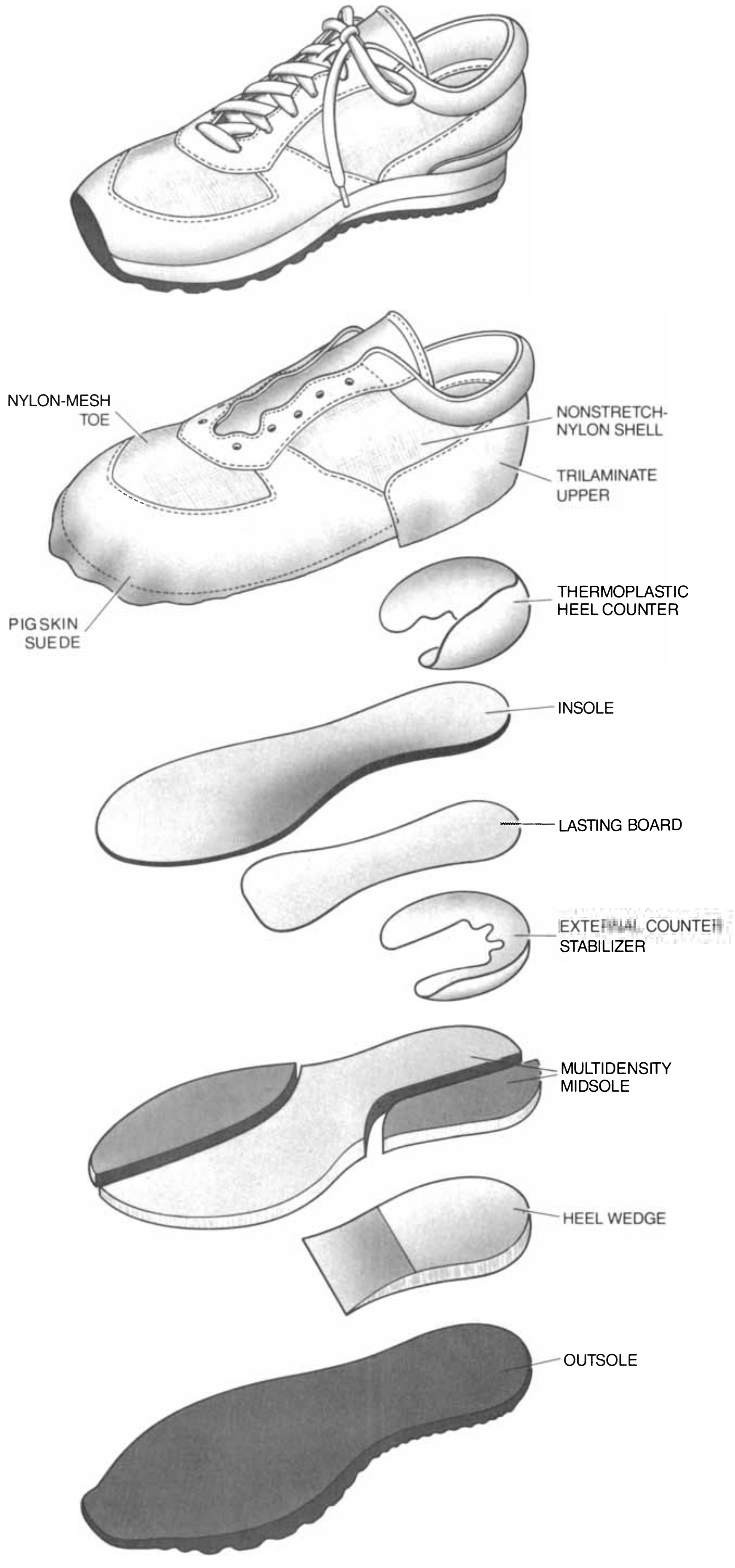
ning shoe's weight of nine ounces, is designed to be as light as possible. It must provide adequate ventilation and endure constant pounding and abrasion. The typical upper made of nylon mesh and a composite material serves the purpose well.

In seeking to improve running shoes technologists are looking for lightweight materials that absorb shocks adequately. Lighter shoes in general mean a lower expenditure of energy by the runner. Measuring submaximal oxygen consumption of runners, Edward C. Frederick and Jack Daniels of Nike have found that each one-ounce reduction in the weight of a single shoe causes a decrease of .28 percent in the energy required for running. They also found, however, that cushioning affects energy consumption. Paradoxically, a light shoe that has improper cushioning will raise the energy requirement.

Ideal running-shoe materials should be able to absorb shock without breaking down. The foams now employed are compressible; use eventually destroys their resilience. Several shoe manufacturers have experimented with soles that have compressed gas embedded in foam. Unfortunately such "air soles" weigh more than a standard foam sole does.

The shock-absorbing capacity of a substance depends mostly on its thickness. Thicker materials afford greater protection. In theory a material that crushes uniformly and gives a constant rate of deceleration through the distance of crushing will result in a minimum force of contact. Such a material, however, would absorb all the energy, returning none to the foot. The result would be low running efficiency. On the other hand, a perfectly elastic spring would return all the energy to the foot but would in theory cause twice the maximum force of a crushable material. Real materials fall somewhere between these limits, returning approximately 40 percent of the energy to the foot.

A study by Richard Bunch of Converse Inc. and his colleagues seeks to tune the response of the shoe to the runner's stride so that a maximum amount of energy is returned to the foot. The trick is to make the com-

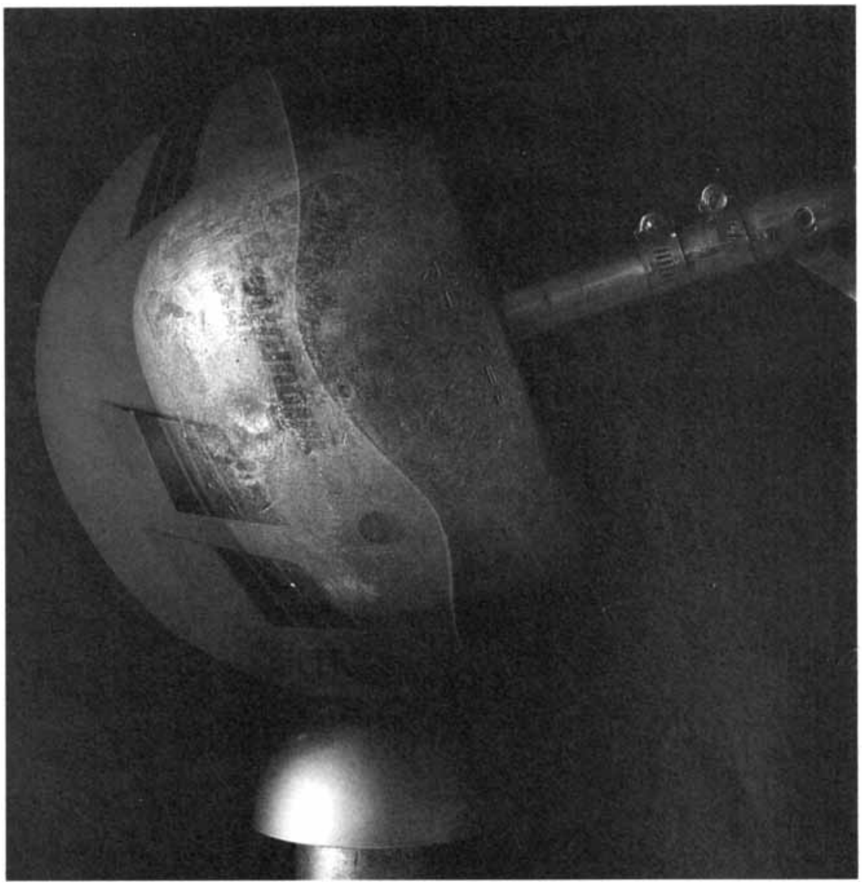


RUNNING SHOE designed for distance running is shown complete at the top and then broken down into its major components. The construction is intended to absorb shock, support the arch and prevent the foot from pronating, or rolling inward, excessively after it makes contact with the ground. Pronation is thought to cause knee injuries.

pressed material of the sole rebound at the same rate as the foot when the foot leaves the ground. The principle has been applied to the design of running tracks. Because runners differ from one another in weight and in weight distribution and have widely varying foot-strike patterns; however, the concept would seem to require custom-made shoes and would therefore be applicable only to the manufacture of footgear for outstanding athletes.

An intriguing thought is that runners could run without shoes on the compliant artificial materials now found widely on outdoor tracks. Nothing is lighter than the bare foot, and it does not pronate. On hard surfaces, however, shoes are essential.

The helmet is probably the most advanced piece of protective equipment in sports. The first organized effort to test helmets scientifically in the U.S. was made by the Snell Memorial Foundation in Wakefield, R.I. It is named for Peter Snell, a racing-car driver who died of head injuries sustained during a race. George G. Snively and Clinton O. Chichester, who with Snell had owned the car in which the driver was injured, set up the foundation in 1957. The foundation devised testing methods and standards for motorcycle and automobile helmets, first publishing them in 1966. The standards were later adopted by automobile- and motorcycle-racing associations throughout the U.S. Snell now publishes standards for ski, bicycle, motor and equestrian helmets.



HEAD FORM is used to test helmets in the laboratory of Bell Helmets, Inc., in Norwalk, Calif. In this double exposure both the form and a helmet under test by the hemispherical anvil can be seen. The test measures the deceleration of the head form. A comparison of that number with the helmetless deceleration shows how much the helmet absorbs the blow.

One standard test ascertains how much a helmet absorbs a blow to the head. In the test the helmet is mounted on a head form, which is then dropped from a height of about three meters onto a flat or hemispherical anvil. An accelerometer mounted at the center of gravity of the assembly measures the peak negative acceleration (deceleration) of the head form on impact. Work in physiology has established that concussion and other compressive injuries to the head will be minimized if the measured peak instantaneous deceleration is less than 300 *g* (300 times normal gravity). The Snell standard for allowable peak deceleration is 285 *g;* most automobile and motorcycle helmets will keep deceleration below that. Moreover, because the metal head form is not a replica of the more compliant human head, the peak decelerations seen in the Snell tests probably represent actual decelerations to a human head of less than 200 *g*.

A helmet is inadequate if it does not stay on the head and if it can be broken or punctured. Additional tests therefore measure the strength of the chin strap and the ability of the helmet's outer shell to resist penetration by a sharp object. Usually the outer shell is made of fiberglass or injection-molded plastic. Fiberglass is the stronger and more durable of the two materials.

Most modern helmets for the motor sports have a crushable expanded-polystyrene liner or a flexible or rigid polyurethane foam to absorb shock. As in shoes, crushable materials are ideal energy absorbers. In the motor and equestrian sports and in skiing, cycling and baseball a helmet can be designed for a single high-energy impact, so that a crushable liner is suitable. In contact sports such as football and hockey the helmet must provide protection against many impacts; the liner must be made of resilient material.

In many sports tradition and the thought that the helmet may impede high performance seem to have more influence on the type of helmet than safety does. Standards for the batting helmet in baseball require that it protect the batter from a fastball thrown at 60 m.p.h., yet many pitchers even at the high school level throw at higher speeds. Since the type of impact is fully predictable, it is well within the capability of modern technology to develop a helmet that could protect a player against today's fastball.

A similar situation exists in cycling. Tradition is so strong that many riders either do not wear a helmet or put on an inadequate leather "hair net" that does not meet the Snell standards or those of the American National Standards Institute. Since racing speeds are higher than 25 m.p.h., contact with the pavement in a crash can cause death or serious injury. A test in which a commonly used leather helmet is exposed to a two-meter drop onto a flat anvil suggests that the head of a cyclist wearing the gear could be exposed to an impact shock greater than 700 *g*. Several manufacturers produce bicycle helmets that give far better protection: their products hold the impact shock in the same test to between 170 and 270 *g*.

Head protection in football has been the subject of study over the past 15 years by Voigt R. Hodgson of the Wayne State University School of Medicine and his colleagues. Their work led to the publication of standards in 1973 by the National Operating Committee on Standards for Athletic Equipment. Somewhat later, rule changes prohibited initial contact with the head in blocking and tackling. Since these two events the number of skull fractures, concussions, paralyzing neck injuries and other serious head injuries in football has declined by more than 50 percent.

Studies at Wayne State and elsewhere have made use of cadavers to measure the severity of impact required to fracture the skull or cause neck injuries. In later tests pressured dye in the cranial arteries of cadavers showed how blood vessels rupture when the head receives a blow. The tests revealed that a wide range of shock intensity led to head injury; the differences arose from differences in bone strength and vascular condition.

Most of the serious head injuries did not occur until peak accelerations of from 100 to 200 *g* were reached. The longer the duration of high acceleration, the more severe the injury. From this information a head-injury criterion was devised to define the danger zone of acceleration.

Hodgson and his associates developed a head model to test impact accurately. The model has the same shape, weight, mass distribution and dynamic response as the human cadaver head. This model in drop tests served to measure the effectiveness of football helmets. Now all players of organized football in the U.S. must wear helmets that pass the resulting standards. They are based on a calculation of acceleration v. time.

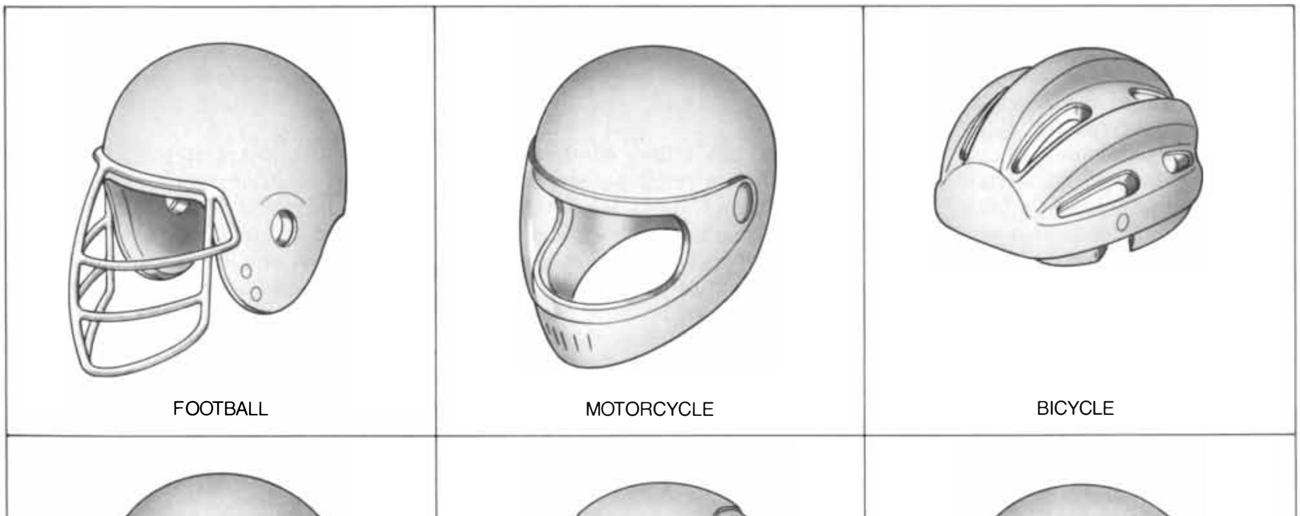
Even the most carefully designed helmet cannot prevent all head injuries. For example, twisting and other forms of rotational acceleration often cause the rupture of blood vessels in the head and other serious damage.

A sport where head protection could be of benefit is boxing. Hodgson is studying the interaction of gloves and head protectors to devise improved helmets for amateur boxers. Perhaps boxing may someday resemble fencing in that a potentially deadly sport is made safe and winning or losing is judged solely on technique and points.

Unfortunately athletes are often reluctant to wear protective gear because of the weight and discomfort or because they assume that its use reflects adversely on their physical courage. Objections to weight and discomfort could be overcome by reducing the weight of protective gear. For example, a goalie in hockey could move faster and make more saves if he wore light padding that gave as much protection as the equipment that is now standard. This is an area that would benefit from further research.

Several major forces are promoting rapid change in sports equipment. The first one is money. The industry making and selling sports and athletic apparel in the U.S. has annual sales that approach $4 billion. Sporting-goods firms compete briskly to sign professional and amateur athletes to contracts binding them to the exclusive use of certain products. The idea is that winning athletes will help the sales of the equipment they use. This strategy creates intense pressure to improve equipment to give sponsored athletes an advantage over their competitors.

Another force for change comes from governments and national athletic organizations, which believe their reputations can be enhanced by victories in international competition. Finally, coaches and athletes constantly call for better equipment. In combination these forces have made technology an increasingly visible and important partner in the design of athletic clothing and equipment.



ARRAY OF HELMETS indicates how the design differs according to the sport. Each helmet except those for football and hockey is designed to protect against a single high-energy impact. The helmets for football and hockey are designed to absorb multiple impacts.

GM

Out here, you've escaped all life's commitments. Except one.

The day's been filled with them. But now you're on your own time. No more interruptions. All commitments temporarily on hold. Except one. Exclusively from GMC Truck.

It was the commitment that began even before you drove your new truck off the GMC lot with a thorough vehicle inspection. It was the satisfaction in seeing a gas tank registering "full" without ever having touched a gas pump nozzle. The security in knowing that a complimentary 1,000-mile inspection followed by a 3,000-mile maintenance check, including free oil and oil-filter change, made the choice to own a GMC so easy.

But GMC's commitment means much more than a few extra conveniences you can actually see and feel. It's the peace of mind you really have knowing that no matter where your GMC truck takes you, the commitment follows.

And that's more than just commitment.
That's Commitment Plus.

GMC
TRUCK
A truck you can live with

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

THE AMATEUR SCIENTIST

Methods and optics of perceiving color in a black-and-white grating

by Jearl Walker

In 1965 Celeste McCollough of Oberlin College reported a puzzling phenomenon: a situation in which the human visual system imposes color on a black-and-white grating. You first study a grating of black stripes interspaced with a color. Later you look at a black-and-white grating identical in orientation and spacing with the first one. The grating's white stripes seem to be tinted with the complement of the color in the first grating.

The apparent color is surprising for several reasons. It appears only if the second grating has the same orientation and spacing as the first one. Although the apparent color may grow fainter with time, it appears even if the observer delays viewing the second grating for hours, days or weeks. (Its strength after a delay depends partly on the dietary and sleeping habits of the observer.)

A person who has normal color vision should be able to see the McCollough coloration by viewing the gratings on the next few pages. Begin with the colored gratings on the opposite page. Note that the black stripes of the horizontal gratings are interspaced with green and those of the vertical gratings are interspaced with magenta. Be sure the illustration is well illuminated. Do not fix your gaze but shift it so that you see each differently colored region for about the same amount of time. After five minutes or more examine the colorless gratings on page 115. The demonstration works better if the illumination is now low. You should perceive faint, unsaturated colors superposed on the white stripes.

The colors you perceive are associated with the orientation of the black stripes. Whereas previously the horizontal stripes were interspaced with green, they are now interspaced with magenta. The vertical stripes change color in precisely the opposite way. You can verify the fact that the colors are associated with the orientation of the stripes by rotating the illustration 90 degrees. Again magenta appears with horizontal stripes and green with vertical stripes. The shapes of the regions are not important; the association of the color is with orientation.

One should not mistake the McCollough effect for the phenomenon known as a negative afterimage. The color in that phenomenon is fleeting compared with the McCollough coloration. To demonstrate a negative afterimage fix your gaze on a field of green color for about five minutes and then on a white, featureless surface. For a short time you will perceive magenta. If you delay viewing the white surface by 10 minutes or more, the afterimage fails to appear.

When you look in a normal way at a white surface, all the color sensors in the retina send signals of color to the brain. Somewhere along the way the signals are analyzed according to pairs of complementary colors that compete. If you see equal amounts of the colors in a pair, they result in a perception of no color. For example, green and magenta compete. If equally strong green light and magenta light reach the eye, the competition is a draw and you perceive white, that is, a colorless illumination.

The normal afterimage is often attributed to fatigue of some of the retinal color sensors. For example, if you view green, the sensors responsible for sending that signal to the brain come to be less responsive. Suppose you view a white surface just after the green sensors become fatigued. Although equally bright green and magenta enter the eye, the weak response by the green sensors allows the magenta to prevail. You perceive magenta in the white region. When the fatigue wears off, the competition between the two colors is again a draw and you see white.

The McCollough coloration also differs from the normal afterimage in that drugs can alter its duration, suggesting it depends on the neurotransmitters in the visual pathway that extends from the retina to the brain. C. C. D. Shute of the University of Cambridge reported that caffeine accelerates the decay of the effect whereas fresh ginseng tea (not the instant variety) delays it. Bamidele O. Amure of the University of Cambridge reported that nicotine prolongs the effect. In addition D. M. MacKay and Valerie MacKay of the University of Keele in England demonstrated that the decay can be postponed by sleep and that the strength of the coloration can depend greatly on how well the observer sleeps before looking at the colored grating.

I did experiments with the McCollough effect by preparing several gratings of black stripes and photocopying them on a machine that retained the solid black character of the stripes. Next I colored the spaces between the stripes. Most often I tested a grating of horizontal black lines interspaced with green. I looked at the grating for about five minutes and then waited five minutes before looking at a colorless grating. The delay ensured that coloring due to a negative afterimage would not appear. The afterimage was normally not a problem anyway since I did not fix my gaze. I had more trouble with the need to allow the McCollough coloration to decay completely between experiments. On occasion the decay took as long as a day.

Does the McCollough effect require straight edges? I made a set of gratings identical with the previous ones except that the black stripes were replaced by rows of small black dots. No McCollough coloration appeared. I replaced the dots with rows of solid black circles but again found no McCollough coloration.

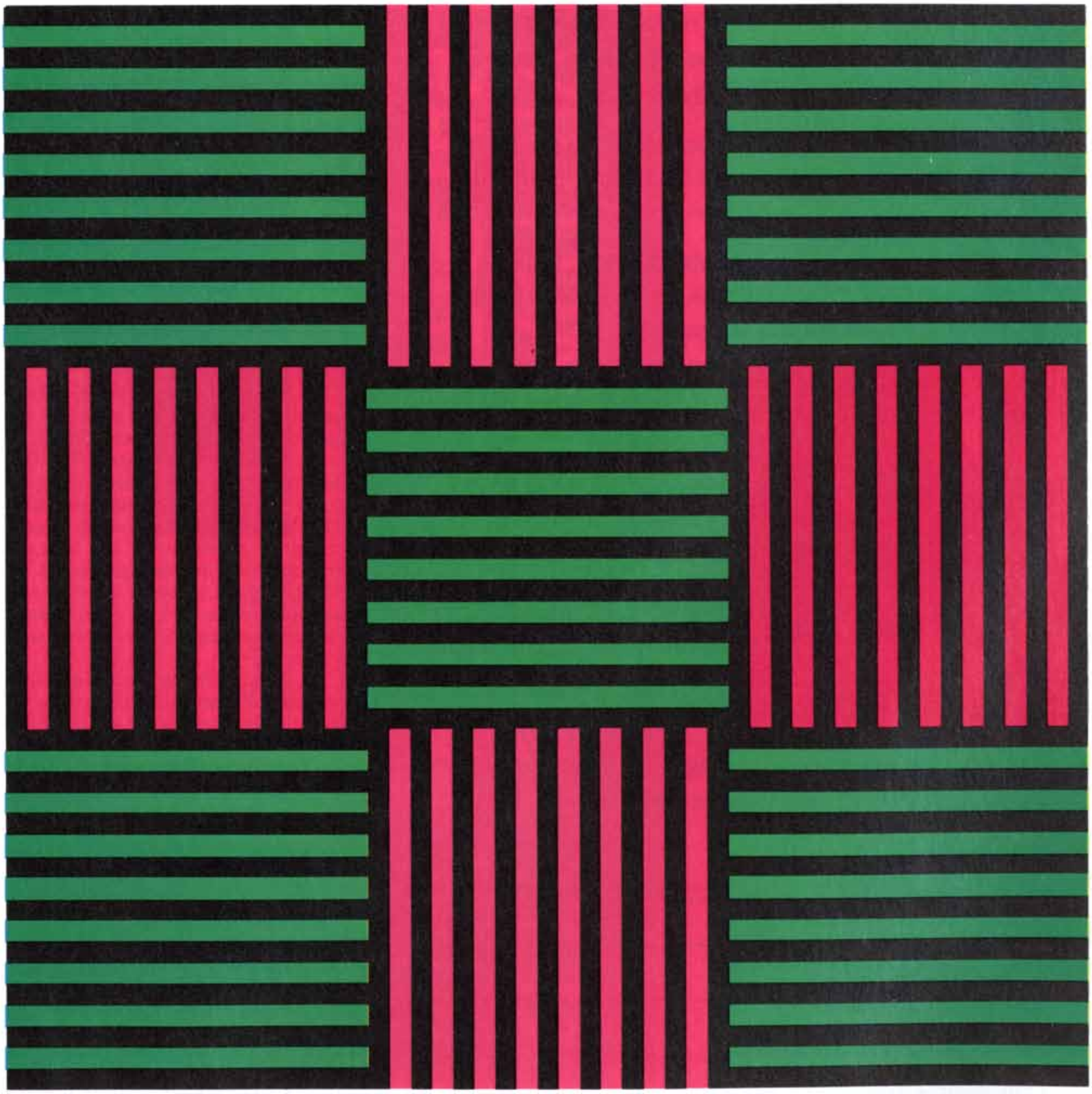
I thought perhaps the spacing of the grating altered the coloration. Conditioning my eyes to the original green-and-black grating, I then looked at a colorless grating that had the same orientation but on which the black-and-white stripes were scaled down to about half the original size. No coloration appeared.

I wondered how much of the field of view must be occupied by the original grating to condition the eyes properly for the color effect. Covering most of the original grating with opaque paper, I gazed at the small amount still visible and then examined the colorless grating without covering any of it. No coloration was present. Apparently one must see enough of the original grating to recognize it as a grating before the McCollough effect will work.

Next I reversed the procedure, looking at the full original grating and then examining only a small part of the colorless one, most of which was covered with opaque paper. The faint magenta appeared. Indeed, it seemed to be there even when I covered up all but a small segment of one or two black stripes. Some students of the McCollough effect have suggested that if the stripes of a grating are widely spaced in one's view, the visual system focuses on the edges of the stripes. If instead the stripes are closely spaced in one's view, the system focuses on the periodicity of the stripes. When I examined the small part of the colorless grating, the McCollough coloration was probably associated with the edges rather than with the periodicity.

In 1974 Charles F. Stromeyer III of Stanford University reported that the McCollough coloration is perceived better if the colorless grating is examined in dim light. The coloration can be seen even when the illumination is so low that the retinal cones, which are responsible for color information, do not function. I tested these results by conditioning my eyes to the black-and-green grating in bright light and then dimming the illumination. As my eyes adjusted to the dim light the magenta coloration in the colorless grating became more pronounced even though I could not see color anywhere else in the room.

Stromeyer also reported that the spacing of the grating can determine the McCollough coloration. An observer first looks at a green-and-black grating with wide spacing. Then he views a closely spaced magenta-and-black grating. The orientations of the two gratings are identical. Some time later he looks at colorless gratings that have the same orientation as the first two. If a colorless grating has the same spacing as the first one, he perceives magenta. If instead it has the same spacing as the second one, he perceives green. Somehow the observer has stored color along with the information about spacing.



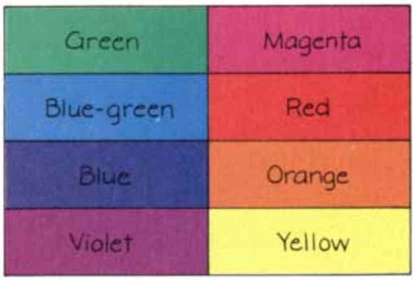
Gratings that condition the eye for the McCollough effect

How is the McCollough effect produced? No one knows in detail, but a few elementary models have been suggested. One of them, which has been put forward by a number of students of the visual system, is shown in the upper illustration on page 118. The model is crude in that details of how it works are not known. At the left is a section of retina illuminated with a horizontal grating of black-and-green stripes. The circuitry at the right represents the early stages of vision. Signals move toward the right to reach the higher levels of processing by the brain. Keep in mind that the illustration is only representative. I do not know where the analyzing sections of the visual system exist, if indeed they can even be localized. I also am uncertain about how they function.

Signals relating to the color of the green stripes are initiated by cone photoreceptors in the retina. The cones come in three types, each of which is sensitive to a different part of the visible spectrum. Once the cones are excited their signals are compared and other signals are relayed deeper into the visual system. These new signals indicate luminosity and the relative strengths of red v. green and of yellow v. blue. At some later stage, which is probably within the brain, the new signals are compared in terms of the competing complementary colors.

For example, the colors green and magenta compete in this later stage of analysis. Consider the competition in terms of numbers assigned to the signal strengths: green positive, magenta negative and an even mixture zero. When the eye sees green, the signal strength might be, say, +200. When it sees magenta, the strength might be −200. If equal amounts of green and magenta are detected (as when you see white), the strength of the color signal is zero.

The edges of the stripes are detected by groups of photoreceptors lying at or near the place where the image of the edge falls on the retina. The information from many edge detectors feeds into a grating detector, which sends



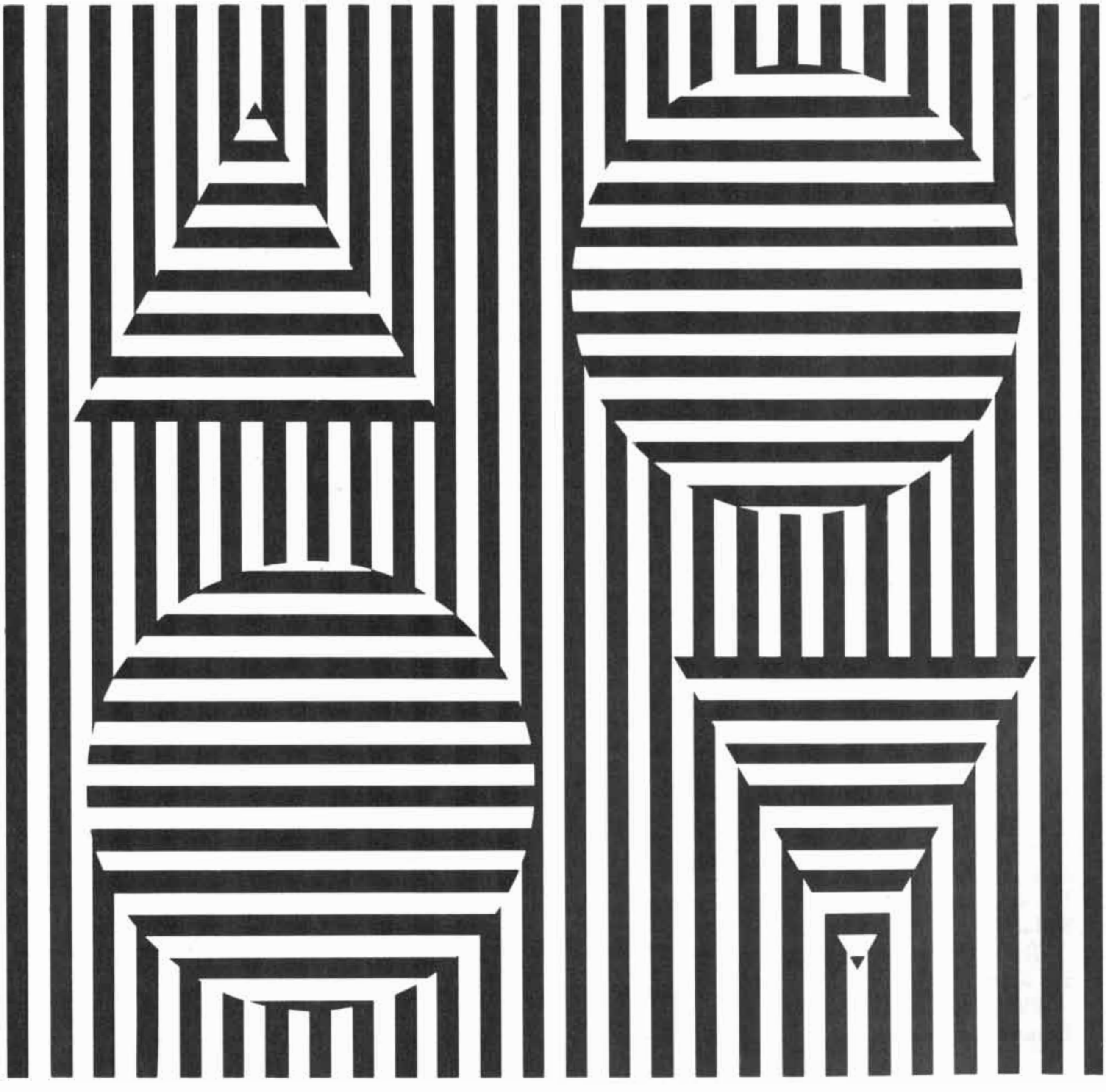
Complementary colors

deeper into the visual system a signal that a grating of a certain orientation and a certain periodicity is being viewed. No one is certain about how the grating detector works. Assume that it recognizes a grating by comparing it with some standard gratings. Once a match is made a signal is sent along an output line. The output line chosen depends on which of the standard gratings matches the one being viewed. If a horizontal grating of a certain spacing is viewed, a signal is sent along one of the output lines. If the grating is rotated to be vertical, the signal is sent along another line. If the spacing of the grating is varied enough to make a match with a different standard grating necessary, the signal is sent along a third line.

The McCollough effect might be explained in terms of this model. As you view a horizontal grating of black-and-green stripes, strong signals are sent along two lines. One signal concerns the green of the light and the other signal concerns the orientation and spacing of the grating. An inhibiting interconnection begins to build between the two lines, somewhat weakening the signal of green. The inhibition is not large enough for you to notice it in normal room light. Although the strength of the interconnection builds up within five minutes or so, it may last for hours or even longer.

Suppose you look later at a colorless grating that has the same orientation and spacing as the first one. On the color line the signal strength is zero because the grating has white stripes. If the color signal continued to be zero, the perception would be of white. The inhibition imposed by the interconnection, however, makes the strength of the signal negative. The color signal then registers magenta, and you perceive magenta on the colorless grating.

Inhibition is imposed on the color line only if the grating detector actuates the output line that has the



Gratings that produce the McCollough effect strengthened interconnection. If some other output line is actuated, the color line is unaltered. Suppose you rotate the grating until it is vertical. The signal from the grating detector moves along another output line, the color line is not inhibited and the McCollough coloration disappears.

Although this model fits many of the experiments done with the McCollough effect, it is unsatisfactory. Why should the output line of a grating detector inhibit the color line? I am also troubled by the following experiment. If the model is correct, the inhibiting interconnection should be built up if you shift your gaze over a colorless, horizontal grating alongside a green region. Yet such an arrangement actually fails to produce a McCollough coloration.

For the fun of it I devised a different model for the McCollough effect [*see lower illustration on page 118*]. I do not know if this model has been studied before. It differs from the preceding one in that the grating detector is itself color-sensitive, sending a color signal into the analyzer that compares complementary colors. The analyzer sums the signal from the grating detector and the principal color signal coming directly from the retina. The strength of the color signal from the grating detector depends on the recognition of the grating.

The color sensitivity of the grating detector may be an unavoidable consequence of the fact that it and the edge detectors analyze signals from color-sensitive retinal cells. In part the detection of an edge may depend on the contrast of color on its two sides. Hence a signal from the grating detector about color is not surprising.

Suppose you view a horizontal grating of black-and-green stripes. After making a match with a standard grating, the grating detector sends a signal out along an output line concerning the orientation and periodicity of the grating. Because of the match, the detector sends to the color analyzer a signal about the green in the grating. Suppose the direct signal is +200 units and the green signal from the grating detector is +10. The signal that emerges from the color analyzer is +210 units, which is perceived as green.

As you continue to examine the grating, the detector grows less responsive to both the grating and its color. After a while the detector begins to reduce its signal about the grating characteristics and you get a poorer perception of the grating. The detector also begins to inhibit the signal of green, sending instead a signal of magenta to the color analyzer. Suppose the strength of the magenta signal is −10 units. If the direct signal of green is again +200, the signal emerging from the color analyzer is +190: a slightly weaker green than before.

When you replace the colored grating with a colorless one of the same spacing and orientation, the grating detector recognizes the grating and again sends a signal of magenta to the color analyzer because of its inhibition to green. Note that at this point the color from the grating detector is triggered by the recognition of the grating, not by the strength of any color signal arriving from the edge detectors. If the magenta signal has a strength of −10 and the direct color signal is zero (because of the whiteness of the stripe being viewed), the color analyzer sends out a signal of −10: a faint magenta. Hence when you view the colorless grating, you perceive a faint magenta superposed on the white stripes. This coloration is the McCollough effect.

What if you look at the colorless grating in dim light? The coloration is more prominent not because of greater signal strength but because the illumination from the white stripes is weaker. You can then pick out the magenta coloring better. The coloration is still apparent even if the light is so dim that the cones no longer function. In such light the visual cells called rods provide vision. They send no color information, but they do serve to detect edges and thus feed a signal to the grating detector. Since the grating detector is still fatigued because you had looked at a grating earlier, the signal it sends to the color analyzer is still −10 units of magenta. You still perceive the coloration.

From the model I figured that under the proper lighting conditions a signal of magenta from a fatigued grating detector might cancel a small, direct signal of green. For example, if the magenta signal is −10 and the direct green signal is +10, they sum in the color analyzer as zero. A green stripe in the grating would then appear to be gray: it is both colorless and dim. Does this graying actually take place? If it does, would the green reappear if I rotated the grating so that the stripes were vertical and the grating detector therefore was forced to carry out a new task of recognition? Suppose an isolated green mark lay in another part of my field of view. Would I still see it as green even if the green of the grating had turned gray?

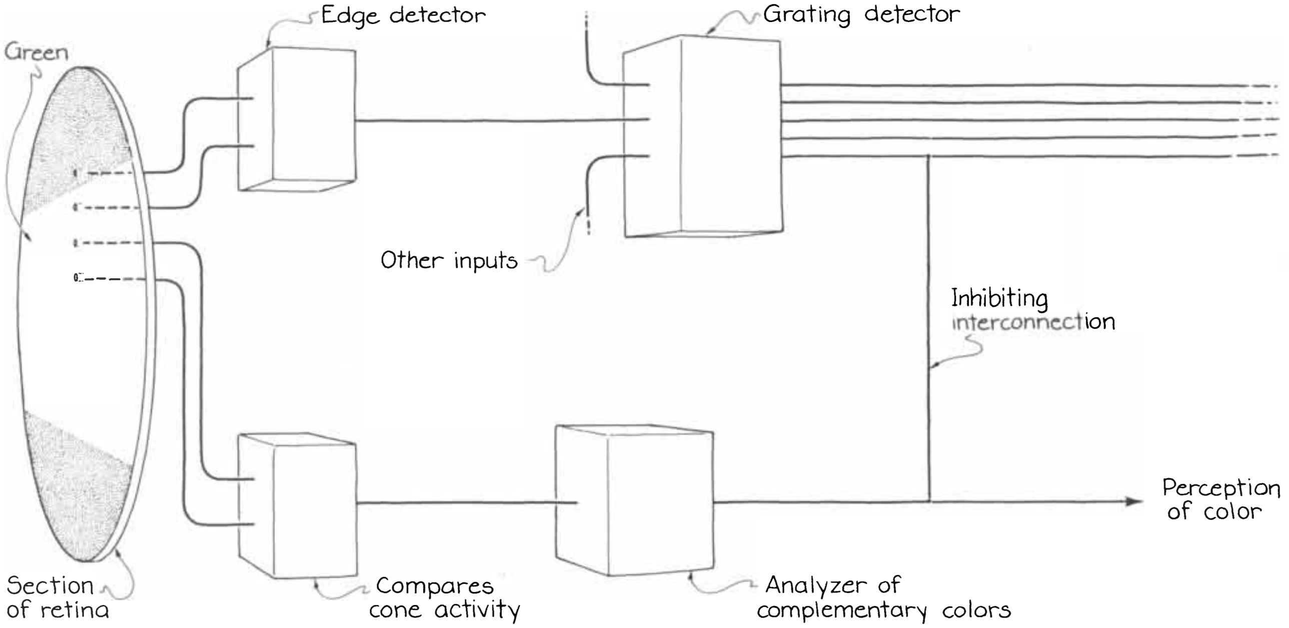
I prepared for the experiment by adding some extra marks just to one side of the usual green-and-black grating. Some of the marks were black and some were green. Some were small and

Dance has a new partner.

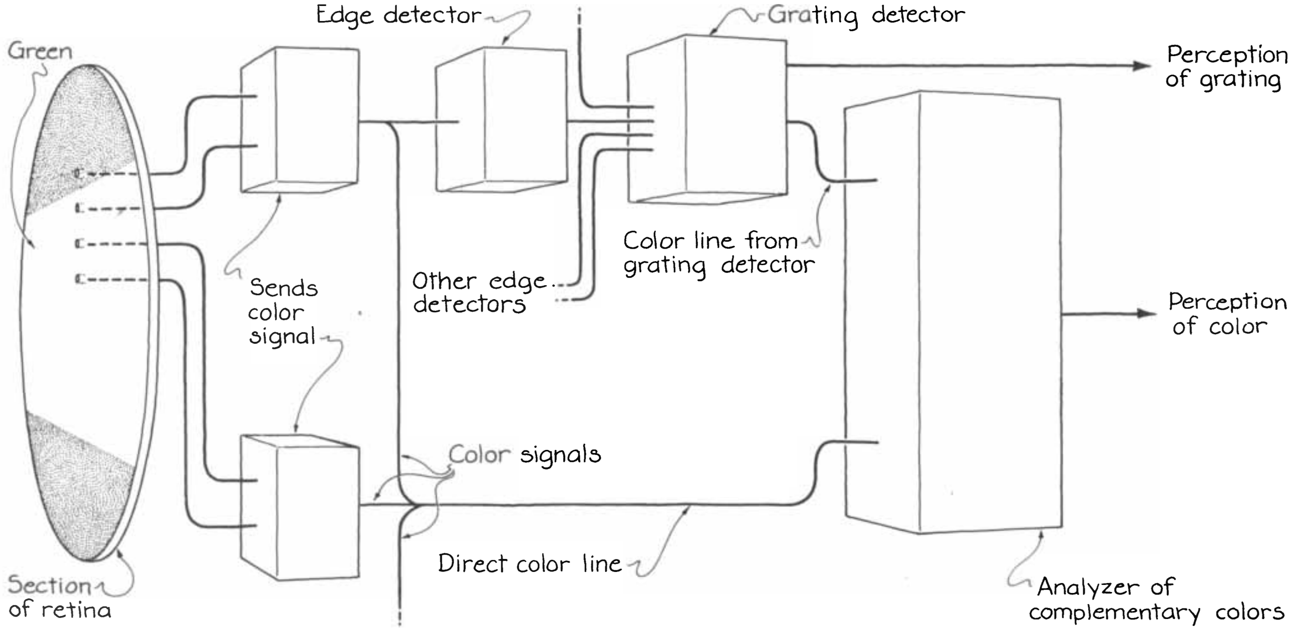
others were stripes as long and wide as those in the grating. I lowered the room lights, waited for 15 minutes so that my eyes would adjust to the illumination and then examined the grating for another 20 minutes.

Although I could initially distinguish the green in the grating, the color soon began to fade into gray. The grating itself became harder to perceive. At this point the grating detector in my visual system must have become less responsive and its color sensitivity must have switched to an inhibition of green. Still, I wondered whether perhaps the gray resulted from a decrease in the direct signal of green because the cones were becoming fatigued by the color. I checked this possibility by shifting my gaze to the isolated marks of green on the paper. Their green was still perceptible. I also checked by rotating the grating until the stripes were vertical. The green of the stripes immediately reappeared. For these reasons the gray of the stripes in the initial orientation seems to be due to the inhibition of green within the grating detector.

Many additional experiments with the McCollough effect have been described in published work. Perhaps you can devise some of your own and can construct a better model than the ones I have considered. If you can, I should like to hear about your work. I would be particularly interested in experiments that disprove my model of a color-sensitive grating detector.

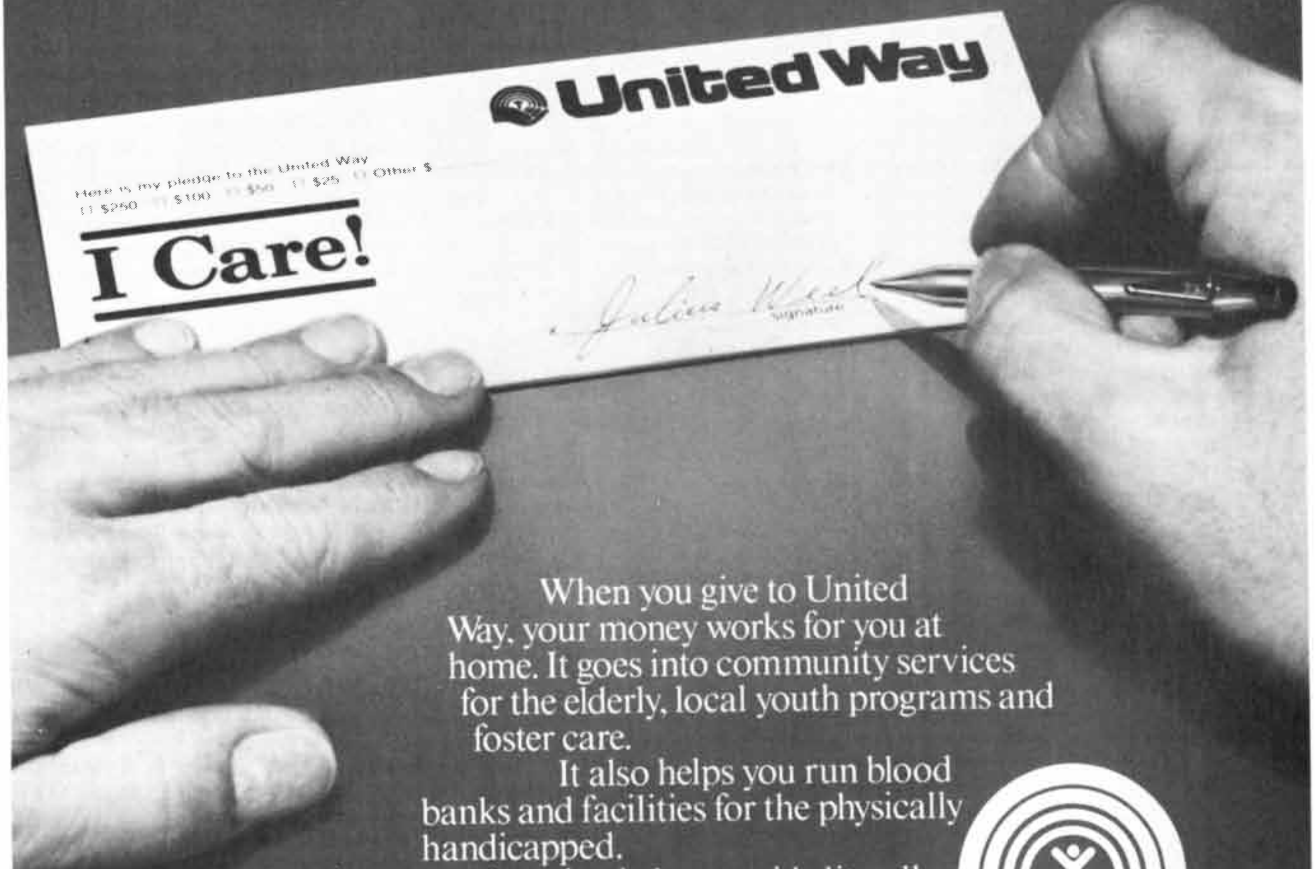


A model of the inhibition of a color signal



A model embodying a color-sensitive grating detector

BUILD A BETTER COMMUNITY WITH YOUR BARE HANDS.


United Way
Here is my pledge to the United Way
□ $250 □ $100 □ $50 □ $25 □ Other $
I Care!
Signature

When you give to United Way, your money works for you at home. It goes into community services for the elderly, local youth programs and foster care.

It also helps you run blood banks and facilities for the physically handicapped.

In fact, your donation helps provide literally hundreds of services that make life a lot better for people in your town.

So when your United Way volunteer comes around, be generous.

A better community is in your hands.

United Way
THANKS TO YOU IT WORKS FOR ALL OF US.

Ad Council

© United Way 1984

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

Artificial Intelligence. Patrick Henry Winston. Addison-Wesley Publishing Co., 1977.

The Mismeasure of Man. Stephen Jay Gould. W. W. Norton & Company, Inc., 1981.

Know Your Own I.Q. H. J. Eysenck. Penguin Books, 1984.

RETHINKING NUCLEAR POWER

Nuclear Power in an Age of Uncertainty. Congress of the U.S., Office of Technology Assessment. U.S. Government Printing Office, 1984.

National Strategies for Nuclear Power Reactor Development. Richard K. Lester, Michael J. Driscoll, Michael W. Golay, David D. Lanning and Lawrence M. Lidsky. Department of Nuclear Engineering, Massachusetts Institute of Technology, March, 1985.

THE EARTH'S MAGNETOTAIL

Transient Phenomena in the Magnetotail and Their Relation to Substorms. E. W. Hones, Jr., in *Space Science Reviews,* Vol. 23, No. 3, pages 393–410; May, 1979.

Majestic Lights. Robert H. Eather. American Geophysical Union, 1980.

The Magnetotail and Substorms. C. T. Russell and R. L. McPherron in *Space Science Reviews,* Vol. 15, Nos. 2/3, pages 205–266; November/December, 1983.

Magnetic Reconnection in Space and Laboratory Plasmas. Edited by Edward W. Hones, Jr. American Geophysical Union, 1984.

THE MOLECULAR GENETICS OF HEMOPHILIA

Blood Coagulation. Craig M. Jackson and Yale Nemerson in *Annual Review of Biochemistry,* Vol. 49, pages 765–811; 1980.

Expression of Active Factor VIII from Recombinant DNA Clones. William I. Wood, Daniel J. Capon, Christian C. Simonsen, Dan L. Eaton, Jane Gitschier, Bruce Keyt, Peter H. Seeburg, Douglas H. Smith, Philip Hollingshead, Karen L. Wion, Eric Delwart, Edward G. D. Truddenham, Gordon A. Vehar and Richard M. Lawn in *Nature,* Vol. 312, No. 5992, pages 330–337; November 22, 1984.

Structure of Human Factor VIII. Gordon A. Vehar, Bruce Keyt, Dan Eaton, Henry Rodriguez, Donogh P. O'Brien, Frances Rotblat, Herman Oppermann, Rodney Keck, William I. Wood, Richard N. Harkins, Edward G. D. Truddenham, Richard M. Lawn and Daniel J. Capon in *Nature,* Vol. 312, No. 5992, pages 337–342; November 22, 1984.

Detection and Sequence of Mutations in the Factor VIII Gene of Haemophiliacs. Jane Gitschier, William I. Wood, Edward G. D. Truddenham, Marc A. Shuman, Therese M. Goralka, Ellson Y. Chen and Richard M. Lawn in *Nature,* Vol. 315, No. 6018, pages 427–430; May 30, 1985.

THE SUPERCONDUCTING SUPERCOLLIDER

The Next Generation of Particle Accelerators. Robert R. Wilson in *Scientific American,* Vol. 242, No. 1, pages 42–57; January, 1980.

The Discovery of Subatomic Particles. Steven Weinberg. W. H. Freeman and Company, 1983.

Concepts of Particle Physics. K. Gottfried and V. Weisskopf. Oxford University Press, 1984.

Superconducting Magnets for Particle Accelerators. R. Palmer and A. Tollestrup in *Annual Review of Nuclear and Particle Science,* Vol. 34, pages 247–284; 1984.

Supercollider Physics. E. Eichten, I. Hinchliffe, K. Lane and C. Quigg in *Reviews of Modern Physics,* Vol. 56, No. 4, pages 579–707; October, 1984.

The Tevatron. H. T. Edwards in *Annual Review of Nuclear and Particle Science,* Vol. 35, pages 605–660; 1985.

COMPUTER-SIMULATED PLANT EVOLUTION

Apparent Changes in the Diversity of Fossil Plants: A Preliminary Assessment. Karl J. Niklas, Bruce H. Tiffney and Andrew H. Knoll in *Evolutionary Biology,* Vol. 12, pages 1–89; 1980.

Paleobotany and the Evolution of Plants. Wilson N. Stewart. Cambridge University Press, 1983.

Mechanical and Photosynthetic Constraints on the Evolution of Plant Shape. Karl J. Niklas and Vincent Kerchner in *Paleobiology,* Vol. 10, No. 1, pages 79–101; Winter, 1984.

MENTAL IMAGERY AND THE VISUAL SYSTEM

Image and Mind. Stephen Michael Kosslyn. Harvard University Press, 1980.

Levels of Equivalence in Imagery and Perception. Ronald A. Finke in *Psychological Review,* Vol. 87, No. 2, pages 113–132; March, 1980.

Mental Imagery and the Third Dimension. Steven Pinker in *Journal of Experimental Psychology: General,* Vol. 109, No. 3, pages 354–371; September, 1980.

CROP STORAGE AT ASSIROS

Excavations at Assiros, 1975–9: A Settlement Site in Central Macedonia and Its Significance for the Prehistory of South-East Europe. K. A. Wardle in *The Annual of the British School at Athens,* No. 75, pages 229–267; 1980.

A Friend in Need Is a Friend Indeed: Social Storage and the Origins of Social Ranking. P. Halstead and J. O'Shea in *Ranking, Resource, and Exchange: Aspects of the Archaeology of Early European Society,* edited by Colin Renfrew and Stephen Shennan. Cambridge University Press, 1982.

Ancient Macedonia: The Cultures. K. A. Wardle in *Macedonia—4000 Years of Greek History and Civilisation,* edited by M. Sakellariou. Edotike Athenon, 1982.

Linear B Tablets and the Mycenaean Economy. J. T. Killen in *Linear B: A 1984 Survey,* edited by A. M. Davies and Y. Duhoux. University of Louvain, 1985.

ATHLETIC CLOTHING

The Running Shoe Book. Peter R. Cavanagh. Anderson World Publishers, 1980.

Fast Fashions. Chester R. Kyle in *Bicycling,* Vol. 26, No. 5, pages 118–125; June, 1985.

THE AMATEUR SCIENTIST

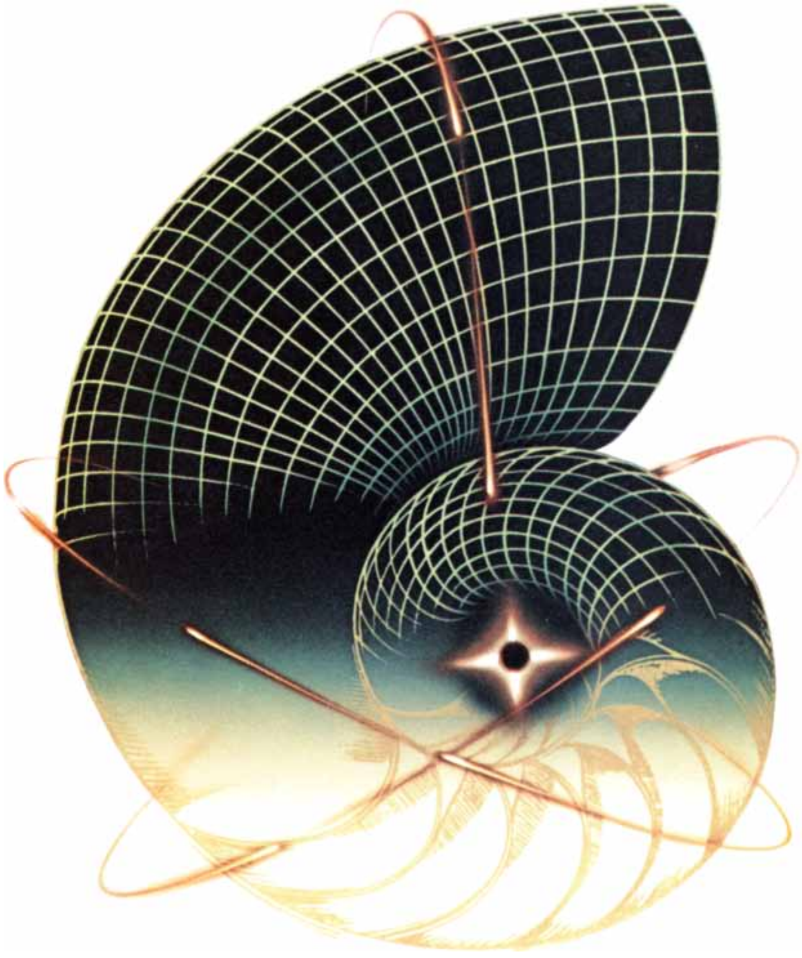
Color Adaptations of Edge-Detectors in the Human Visual System. Celeste McCollough in *Science,* Vol. 149, No. 3688, pages 1115–1116; September 3, 1965.

Form-Specific Colour After Effects in Scotopic Illumination. Charles F. Stromeyer III in *Nature,* Vol. 250, No. 5463, pages 266–268; July 19, 1974.

'ek·sə·ləns

Excellence

• Distinguished by superiority; e.g., 123,000 committed, dedicated employees—highly skilled in complementary technologies.
• A standard, like the quality of products being developed in our five core business areas.
• A surpassing achievement, such as '85 sales exceeding $11 billion.



$$F = FR_1 / \frac{d}{d\alpha} \left[\log \left(T - T_{pv} - T_{os} \right) \right]$$

The Nautilus Configuration: It is nature's inspiration for excellence in science and engineering. It is also a symbol of the commitment of Rockwell employees.

The quest for excellence by our 25,000 engineers, scientists and supporting technical personnel shapes the work of Rockwell's 123,000 employees worldwide. And results in the elegant solutions to customer needs that make Rockwell a leader in five diverse areas of commercial and government business.

Excellence is also a major reason for the ongoing record of financial growth that brought us more than $11 billion in sales and record earnings in 1985.

To learn more about us, write: Rockwell International, Department 815S-3,
600 Grant Street, Pittsburgh, PA 15219.

Rockwell International

...where science gets down to business

Aerospace / Electronics / Automotive
General Industries / A-B Industrial Automation

Fly First Class.
Wild Turkey. It's not the best because it's expensive.
It's expensive because it's the best.
Austin Nichols
WILD
TURKEY
KENTUCKY STRAIGHT BOURBON
WHISKEY
101 PROOF
8
YEARS OLD