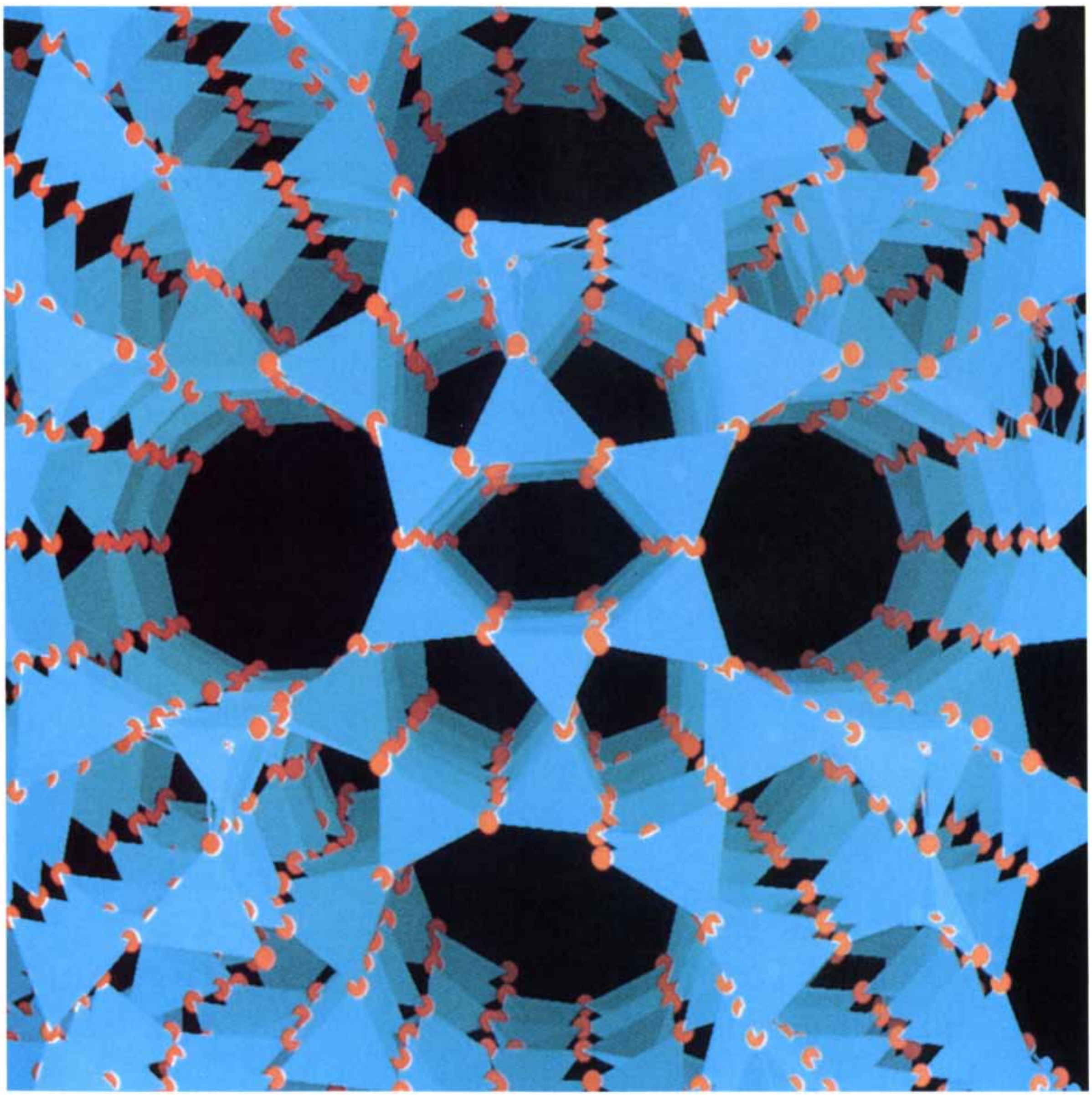
SCIENTIFIC AMERICAN

JULY 1989
$2.95

Mapping the course of memory lane.

Do floods of magma help to build continents?

Spin glasses: simplifying complexity.



***Porous zeolite crystals** perform key functions in industry.*
This one helps to convert wood alcohol into gasoline.

© 1989 SCIENTIFIC AMERICAN, INC

The guys at Pola
to shoot each ot

Picture it. Their data system was good.

But now, with the instant success of their Spectra® camera, good suddenly wasn't good enough.

!@#$%&*( )! Trouble.

So Polaroid called in NYNEX—one of the information industry's favorite troubleshooters.

And NYNEX designed an integrated voice and data network that connects all eleven Polaroid locations.

And makes everything from inventory to finance to technological analysis faster. Easier. And more cost-effective.

There were simply no negatives.

The NYNEX family of companies would like to work with you, too.

© 1989 SCIENTIFIC AMERICAN, INC

roid were ready

ier.

We offer you everything from computer networks and software, to mobile phones, to the services provided by New England Telephone and New York Telephone. And more.

Call us at 1 800 535-1535.

Discovering the answer is NYNEX could be a long-term boon to your business.

We're hardly a flash in the pan.

Need to communicate? Need to compute? The answer is

NYNEX

© 1989 SCIENTIFIC AMERICAN, INC

SCIENTIFIC AMERICAN

July 1989 Volume 261 Number 1

© 1989 SCIENTIFIC AMERICAN, INC

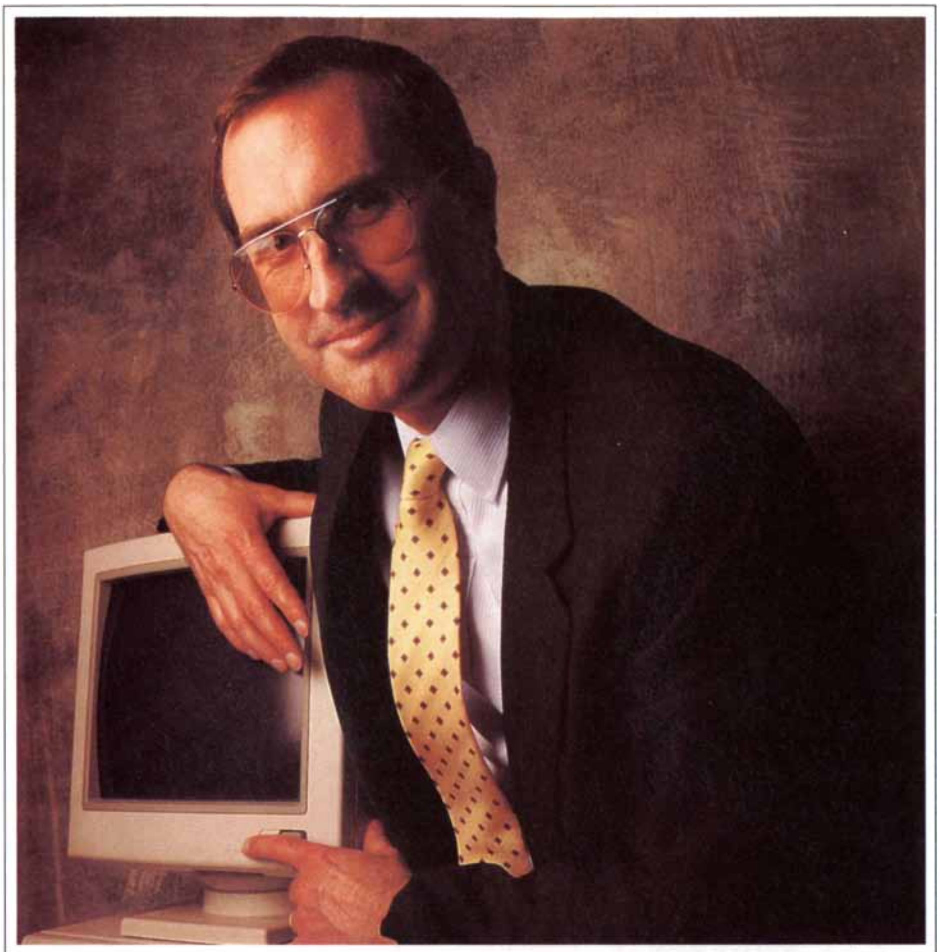
DEPARTMENTS

© 1989 SCIENTIFIC AMERICAN, INC

Ted Standish has a bear of a job.

As Director of Information Resources for the North Atlantic Shaving Group of The Gillette Company, he's been given the task of simplifying and standardizing the business systems of the Company's European and North American divisions.

"With the coming of the European open market in 1992, we want to take a more unified approach to products, as well as to manufacturing and marketing," says Ted.



Ted Standish, Director of Information Resources, North Atlantic Shaving Group of The Gillette Company.

"Before Lotus Agenda, there were three things I always forgot. Tasks I assigned, promises I made, and I can't remember the third."

"This means dealing with hundreds of issues, projects, facts and people throughout the North Atlantic region every day.

"As you might imagine, reams of little pieces of paper with notes, phone numbers, ideas and who-knows-what-else were beginning to take over my office."

Which is why Ted visited his local computer software reseller and bought Lotus® Agenda®, the personal information manager.

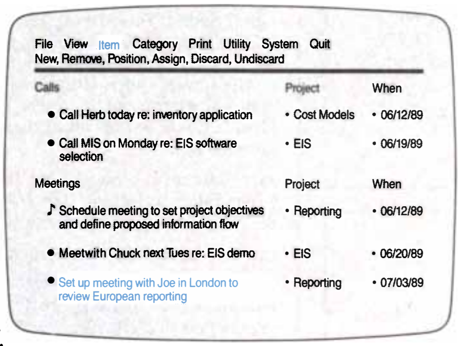
In Agenda, Ted has a tool that's helping him manage people, projects, ideas, plans and information, all with incredible efficiency.

Because Agenda allows him to enter

Agenda is also available in an OS/2® version. If for any reason you're unsatisfied with Agenda, return it with your sales slip within 30 days of purchase and we'll refund your full purchase price.

© 1989 SCIENTIFIC AMERICAN, INC

information into his PC without having to structure it in advance. Agenda automatically arranges the data any way Ted wants so he can understand it better, formulate new ideas better and extract the answers he needs better.



The new Agenda Activities Planner gives you step-by-step instructions that put you in control of your information quickly.

"It lets me focus more on using my information and less on processing it," says Ted.

"For example, I can dump in all my 'to do's' and have Agenda organize, prioritize and categorize them automatically. And whenever I call up a project, person or due date, I'm instantly given all the background information I need to make smart decisions.

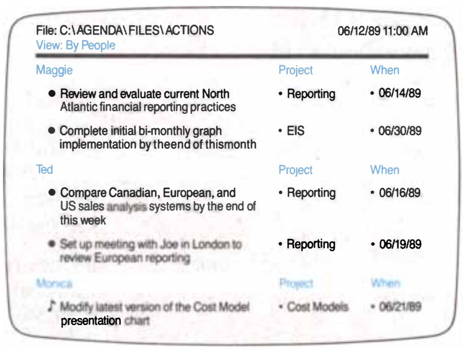
"Now I no longer worry about what I told someone in England, or what somebody else set up in France, or who's in Boston doing whatever.

"Agenda allows me to quickly consolidate all these things in a single place. And discover relationships between people and responsibilities, projects or any other relevant topics that come up.

"Now I can see at a glance who's promised to deliver what and by when. More important, I can keep my own promises."

And Ted isn't the only business person who's come to rely on Lotus Agenda.

The Royal Bank of Canada is using it to monitor international economic and political activities. The accounting and consulting firm of Laventhol & Horwath is using it to track and support its computer systems all over the country. And the CAP International market research company is even using it to predict trends.



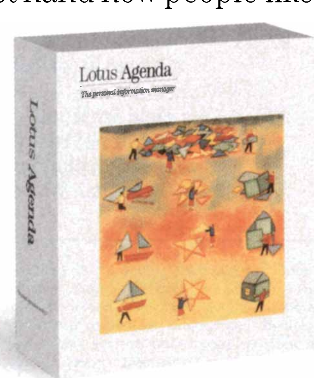
Agenda automatically files items in all relevant categories. Then you can view information from different perspectives.

The list goes on and on. The point is, Agenda can significantly help you, too. Especially now that it comes with an Activities Planner that gives you simple, step-by-step instructions so you can sit down and create your own personalized task manager.

If you'd like to learn even more about how Agenda can put you in control, just send us $10 for an Agenda demo kit. **Call 1-800-345-1043 and ask for demo kit AOP-3058 for 3.5" or AOP-3053 for 5.25"** (of course, we'll refund your $10 when you buy Agenda).

Then you'll see first hand how people like Ted Standish are working better, thinking better and solving problems better.

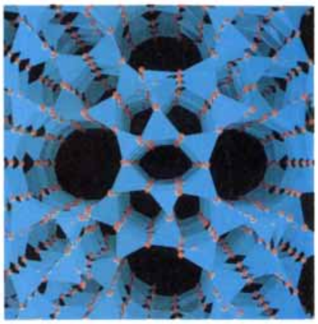
Not to mention remembering better.



Lotus Agenda

© 1989 Lotus Development Corporation. Lotus and Agenda are registered trademarks of Lotus Development Corporation. OS/2 is a registered trademark of International Business Machines Incorporated

© 1989 SCIENTIFIC AMERICAN, INC



THE COVER image is a model of the catalytic zeolite crystal ZSM-5 (see "Synthetic Zeolites," by George T. Kerr, page 100). It is one of many synthetic zeolites important in industry, particularly in petroleum refining. The blue shapes are silicon- or aluminum-centered tetrahedrons whose vertexes are filled by oxygen atoms (*red spheres*); the oxygen atoms are shown disproportionately small. The image was made with Chem-X software, developed and distributed by Chemical Design Ltd., in Oxford, England.

THE ILLUSTRATIONS

Cover image by Chemical Design Ltd.

Page	Source
34-37	Ian Worpole
38	Tass/Sovfoto
39	Johnny Johnson
40	Tass/Sovfoto
43	James L. Olds, National Institute of Neurological and Communicative Disorders and Stroke
44-50	Thomas C. Moore
53	Quesada/Burke
54-57	Hank Iken
63	Keith G. Cox, University of Oxford
64-70	George Retseck
79	National Art Museums of Sweden
80-83	Patricia J. Wynne
85	Artur A. Tishchenko
86-90	George V. Kelvin
92-93	Institut für Urgeschichte, University of Tübingen
94	Field Museum of Natural History, Chicago; photographs by Randall White
95	Institut Royal des Sciences Naturelles, Brussels; photographs by Randall White
96	Institut für Urgeschichte, University of Tübingen (*top*), Musée National de Préhistoire, Les Eyzies-de-Tayac (*bottom*)
97	Musée des Antiquités Nationales, Saint-Germain-en-Laye; Logan Museum of Anthropology, Beloit College; and Musée des Antiquités Nationales; photographs by Randall White
98	Logan Museum of Anthropology; photograph by Randall White
99	Field Museum of Natural History; photographs by S. Varnedoe
101	Chemical Design Ltd.
102-104	Gabor Kiss
105	Mark E. Davis, Virginia Polytechnic Institute and State University
106-109	Michael Goodman
110-113	Clifford A. Pickover, IBM Thomas J. Watson Research Center

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Copyright © 1989 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Subscription rates: one year $27, two years $48, three years $66 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51593.

SCIENTIFIC AMERICAN

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Armand Schwab, Jr., *Managing Editor;* Timothy Appenzeller, *Associate Editor;* Timothy M. Beardsley; John M. Benditt; Laurie Burnham; Elizabeth Corcoran; Gregory R. Greenwell; John Horgan; June Kinoshita; Philip Morrison, *Book Editor;* Tony Rothman; Ricki L. Rusting; Russell Ruthen; Paul Wallich; Karen Wright

ART: Samuel L. Howard, *Art Director;* Murray Greenfield, *Associate Art Director;* Edward Bell, *Assistant Art Director;* Johnny Johnson

COPY: Maria-Christina Keller, *Copy Chief;* Michele S. Moise

PRODUCTION: Richard Sasso, *Vice-President Production and Distribution; Managers:* Carol Eisler, *Manufacturing and Distribution;* Carol Hansen, *Electronic Composition;* Leo J. Petruzzi, *Manufacturing and Makeup;* Carol Albert; Madelyn Keyes; William Sherman; Julio E. Xavier

CIRCULATION: Bob Bruno, *Circulation Director;* Lorraine Terlecki, *Business Manager*

ADVERTISING OFFICES: NEW YORK: Scientific American, 415 Madison Avenue, New York, NY 10017; Robert F. Gregory, *Advertising Manager;* Lisa Carden; Peter Fisch; John Grant; Meryle Lowenthal. CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager;* Litt Clark, *Midwest Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; William F. Moore, *Advertising Manager;* Edward A. Bartley, *Detroit Manager.* WEST COAST: 1650 Veteran Avenue, Suite 101, Los Angeles, CA 90024; Kate Dobson, *Advertising Manager*; Joan Berend. ATLANTA: Quenzer/Stites. CANADA: Fenn Company, Inc. DALLAS: Griffith Group. PRINCETON: William Lieberman, Inc.

ADVERTISING SERVICES: Laura Salant, *Sales Services Director;* Diane Greenberg, *Promotion Manager;* Ethel D. Little, *Advertising Coordinator*

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; GWP, Düsseldorf. HONG KONG/SOUTHEAST ASIA: C. Cheney & Associates. SEOUL: Biscom, Inc. SINGAPORE: Cheney Tan Associates. TOKYO: Nikkei International, Ltd.

ASSOCIATE PUBLISHER/BUSINESS MANAGER: John J. Moeling, Jr.

PRESIDENT OF MAGAZINE DIVISION AND PUBLISHER: Harry Myers

SCIENTIFIC AMERICAN, INC.

415 Madison Avenue
New York, NY 10017
(212) 754-0550

PRESIDENT AND CHIEF EXECUTIVE OFFICER: Claus-Gerhard Firchow

EXECUTIVE COMMITTEE: Claus-G. Firchow; *Executive Vice-President and Chief Financial Officer,* R. Vincent Barger; *Senior Vice-President,* Harry Myers; *Vice-Presidents:* Linda Chaput, Jonathan Piel, Carol Snow

CHAIRMAN OF THE BOARD: Georg-Dieter von Holtzbrinck

CHAIRMAN EMERITUS: Gerard Piel

© 1989 SCIENTIFIC AMERICAN, INC

You can love it without getting your heart broken.

All over the country, people are breaking up with their cars in record numbers. It's in the paper every day. Just check the used-car section.

You could avoid this kind of heartbreak with a little foresight and a trip to your local Saab dealer.

There, you might fall in love with a Saab like the 900 Turbo Convertible pictured here. You could do it with an easy conscience, because unlike most convertibles, ours is a Saab as well.

It wouldn't elate you in the spring, then let you down in the gloom of winter. Like all Saabs, it features front-wheel drive for great traction in any weather. Plus heated front seats and a double-insulated top for comfort, heated glass rear window for visibility and reinforced windshield pillars for safety.

You could drive it all year long and not realize it's a convertible until you dropped the top.

The Saab 900 Turbo Convertible is just one of the ways Saab dealers help keep people from buying the wrong cars–cars that offer fun without practicality or practicality without fun or status at prices that sting long after the honeymoon is over.

Don't buy the wrong car. Don't fall in love with a car you can't live with. Instead, make a date to test drive a Saab.

Your Saab dealer will be more than happy to set you up.

SAAB SCANIA

SAAB
The most intelligent cars ever built.

Saabs are intelligently priced from $16,995 to $32,095. 900 Turbo Convertible shown here: $32,095. MSRP, not including taxes, license, freight, dealer charges or options. Prices subject to change. ©1989 Saab-Scania of America, Inc.

© 1989 SCIENTIFIC AMERICAN, INC

LETTERS

To the Editors:

In his stimulating article "The Electrification of Thunderstorms" [SCIENTIFIC AMERICAN, November, 1988], Earle R. Williams discusses the convection hypothesis. We question his judgment that the coulomb of charge being deposited each second on the moving surfaces of a roiling, poorly conducting cloud probably plays only a minor role in its electrification.

We disagree that the convection hypothesis is "somewhat more complicated" than the precipitation hypothesis. The convection model is complicated by uncertainties concerning how air moves within the cloud, but so are all the precipitation models. Because airspeeds up, down and sideways may be several times the fall speed of precipitation particles, convection dominates most charge-accumulation processes in clouds.

We also disagree with Williams' assessment that the screening layer predicted by the convection model is only a secondary feature of the cloud's electrical structure. Measurements show that it carries a charge comparable to that transferred by an intracloud lightning flash.

In 1960 we carried out experiments in Illinois to see what effects the release of space charge from a high-voltage wire would have on the electrification of fine-weather cumulus clouds. We question Williams' conclusion that our experiments "cannot be said to strongly support the convection hypothesis." On the contrary, we found that a cloud growing from the artificially charged air exhibited electric-field and space-charge perturbations in quantitative agreement with the theory, which predicted that if the cloud had grown to thunderstorm size it would have been capable of causing lightning. Williams' point that the electric fields produced by these small clouds were more than 1,000 times smaller than those believed necessary to produce lightning is of doubtful relevance. Our extrapolation of 10^3 from small cumuli to a thundercloud is probably no more risky than Williams' extrapolation of 10^{15} from laboratory experiments on ice particles.

That "the precipitation model can account for more aspects of cloud electrification than the convection model" is doubtful, for the model described by Williams provides few clues to the following: the fate of the charges released beneath the cloud by the corona and those deposited on the top of the cloud by conduction; why the frequency of lightning increases with cloud height and with the conduction current flowing to the cloud top; how some clouds can make lightning even though they are everywhere warmer than zero degree Celsius; how volcano clouds produce lightning; and how lightning is caused on other planets, where the temperature and composition of clouds are quite different.

An important requirement of any electrification model is that it explain what happens to the charge being deposited on the top of the thundercloud by the current flowing in the global circuit. The assumption that downdrafts carry this charge to lower levels is the primary uncertainty of the convection hypothesis. Recently developed, circularly polarized radar technology may make it possible, by means of radar-reflective chaff particles serving as tracers, to determine whether or not the air in and around the cloud moves in this way.

BERNARD VONNEGUT

State University of New York
at Albany

CHARLES B. MOORE

New Mexico Institute
of Mining and Technology

The author responds:

Bernard Vonnegut and Charles B. Moore (VM) have been major contributors to research in thunderstorm electricity for nearly 40 years, arguing in favor of the convection hypothesis on the basis of both theory and observation. Their infectious enthusiasm was in no small way responsible for drawing me into this field.

In regard to the macroscopic separation of charge (the microphysics of charge transfer in ice-particle collisions is indeed a mystery), I stand by the statement that the convection hypothesis is "somewhat more complicated" than the precipitation hypothesis. No matter how complicated and vigorous the convective motions are in a cloud, an orderly differential motion of particles of different sizes is guaranteed to occur. Furthermore, vertical displacements of selectively charged precipitation particles by only 500 meters can account for the observed dipole structure of thunderclouds, whereas the convection theory requires vertical displacements of space charge by 5,000 meters or more.

The main negative layer in a thundercloud and the upper and lower positive regions have all been shown to be active participants in lightning. The negative screening layer at the cloud top and the corona layer beneath the cloud, which are fundamental to the convection theory and which are seen consistently, have not been shown to be active participants.

Referring to 1960 experiments with small clouds, VM claim their theory predicted that if the clouds had grown to thunderstorm size they would have been capable of causing lightning. Why, then, did no quantitative calculation in support of the idea appear in their *Science* paper? In contrast, calculations based on field data and on laboratory studies of ice-particle collisions have shown that the ice-based precipitation mechanism is capable of causing lightning.

The precipitation-based mechanism I described does provide clues to many of the phenomena VM list. Lightning frequency increases with cloud height, for example, because higher clouds contain more ice particles in the mixed-phase region. The conduction current flowing to the cloud top increases because of the increased lightning frequency: more negative charge must then flow to the cloud top to allow for charge conservation. (The charge-separation process within the cloud is protected from the dissipative effects of conduction by the screening layer.) In a volcano cloud the range of particle sizes and the corresponding range of differential velocities are undoubtedly far greater than during a thunderstorm. Even though the microphysics responsible for the charging of dissimilar particles is still poorly understood, the process may play a principal role here. The other planet for which evidence of lightning is most convincing is Jupiter, and its atmosphere is believed to contain clouds of liquid water and ice.

I agree completely with VM that the ice-based precipitation mechanism offers no clue to the production of lightning by warm clouds. Of the few reports of this phenomenon, however, only one—by Moore, Vonnegut and their colleagues in 1960—is adequately documented.

VM point to observations that any theory of cloud electrification should explain. I think it is important that, in addition to accounting for their own observations (and interpretations), the convection model also explain observations by other competent scientists.

EARLE R. WILLIAMS

© 1989 SCIENTIFIC AMERICAN, INC

Sony just added a startling new dimension to photography.

It's called televison.

With the new Sony Mavica® still video camera, you can now capture color images electronically. Then just connect the camera and multi-featured playback controller* to any TV and see instant magic right on the screen.

Using the same technology Sony invented for photo journalists, the Mavica is an amazing advance that fills the void between still and video cameras. And since this sleek, simple device is as portable as a 35mm camera, photo opportunities will abound.

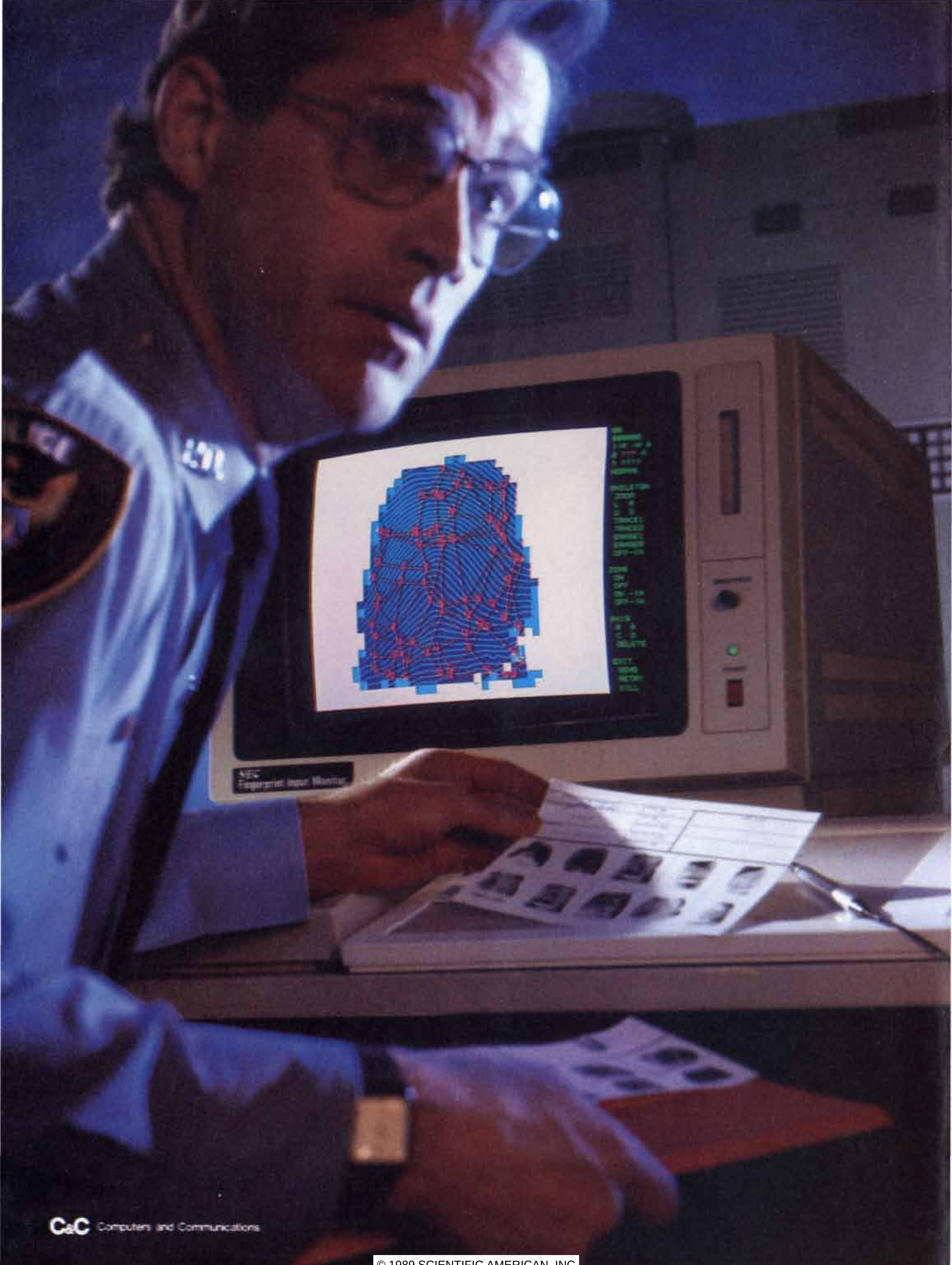
Unlike film cameras, the Mavica records 50 images on a 2″ reusable Mavipak® disk. With no processing and no waiting. Nothing could be easier for people on the go who want to stop and see life one moment at a time.

The new Mavica camera.

Once again, Sony's vision improves your vision.

SONY® Mavica®

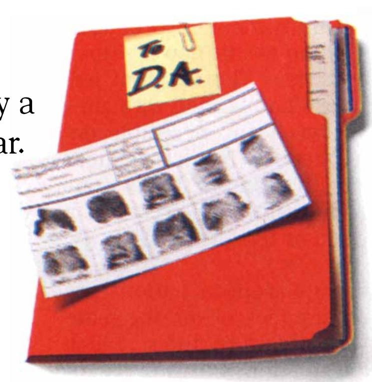
©1989 Sony Corp. of America. Sony, Mavica and Mavipak are registered trademarks of Sony
© 1989 SCIENTIFIC AMERICAN, INC



© 1989 SCIENTIFIC AMERICAN, INC

10:34 AM

NEC helps the Chicago Police solve a fifteen-year-old murder mystery.



An unresolved murder baffled the Chicago Police. It sat in their files year after year with only a single clue: fingerprints lifted from the victim's car.

The prints proved useless—until detectives put NEC's revolutionary Automated Fingerprint Identification System (AFIS) on the case. Within 20 minutes, a suspect who had gotten away with murder for 15 years was identified. He was later arrested.

The quality and reliability of NEC's around-the-clock system is currently at work in police departments all across America. In fact, since its introduction, the system has been instrumental in identifying over 16,000 suspects throughout the U.S.

For years, NEC quality has reached far beyond law enforcement, into every corner of America. You'll find the quality products of NEC in corporations large and small. In public and private institutions. And in homes across the country. Everywhere you look, NEC quality is proving itself. Every minute of every day.

For more proof, call 1-800-338-9549.

NEC

Quality That Proves Itself.

Communications • Computers • Electron Devices • Home Electronics

© 1989 SCIENTIFIC AMERICAN, INC

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JULY, 1939: "INCOMPARABLE PROMISE OR AWFUL THREAT? For years past, scientific writers have been pointing out that if the energy known to be confined within the atom could be practically and commercially released and employed, there would be power in amounts today unknown. The atomic energy of a piece of coal would exceed that of the same coal burned beneath a boiler by a ratio of a whole billion to one. As this is written, physicists are at work on experiments that seem to contain the possibility if not the probability that the billion-to-one power source is at least in view. The uranium isotope 235, when bombarded with slow neutrons, releases inconceivable energy. Simultaneously it releases more neutrons and, if uranium 235 isotope be present in quantity, the result is a chain reaction of terrible power."

"Since most European nations are deficient in supplies of oil, the manufacture of substitute fuels, which permit conservation and storage of gasoline against a possible war, has been forced to a new high since 1937. Wood, charcoal, coal and lignite are being used to make producer gas in generators on many European motor cars. Compressed natural and manufactured illuminating gases are carried in clumsy tanks on some cars. Experiments have been conducted in efforts to utilize acetylene, cracked ammonia and hydrogen. Hydrogenated gasoline—made from either coal or carbon monoxide—is, however, Europe's primary motor fuel."

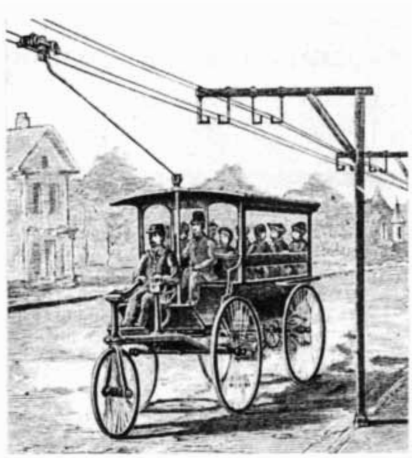
"Among the manufacturing operations performed on aluminum alloys, none is more interesting and spectacular than the impact extrusion process. By this method a small, biscuit-shaped slug of aluminum is transformed in the twinkling of an eye into a toothpaste tube, a flashlight case or any one of several hundred commonplace containers and shapes. A slug of aluminum is placed on the die, the hammer strikes the slug and the high pressure forces the aluminum slug to 'flow' up past the clearance between die and hammer, thus forming a collapsible tube or any other shape determined by the die. The metal is literally 'squirted' into the desired shape."

"What keeps the stars shining? According to Bethe's theory, the most notable achievement of theoretical astrophysics of the past 15 years, it is the transformation of hydrogen into heavier elements. From the mere statement that there is a body having 330,000 times the Earth's mass and containing 35 percent of hydrogen, the theory enables us to predict that it will have the size, brightness and temperature of the Sun. We could do the same for Sirius and for innumerable other stars. Thus the Main Sequence is theoretically explained."

SCIENTIFIC AMERICAN

JULY, 1889: "Thomas A. Edison recently revealed some of the characteristics of his mode of living: 'Yes; I am a hard worker. You know, when I am working on anything I keep at it night and day, sleeping a few hours with my clothes on. I sleep from 1:00 to 6:00 in the morning, and then I jump up and go to work again as fresh as a bird. But I tell you we have lots of fun in the laboratory. Some time ago I had forty-two men working with me on the incandescent lamp in a big building. I hired a German to play an organ for us all night, and we worked by the music. Oh, we enjoy this kind of life!'"

"*Science* describes the experiences of a gentleman who placed himself under the influence of a hasheesh. He smoked it until he felt a profound sense of a well-being and then put the pipe aside. He felt an unusual fullness of mental impressions—enough to fill volumes. He could not write one word without hurrying to the next, his thoughts flowing with enormous rapidity. The few words he did write meant nothing."



Dibble's electrically propelled carriage

"Professor John Trowbridge calls attention to the importance, from an engineering point of view, of making careful photographs of steel and timber at the point of rupture under a breaking load, suggesting that in this way we may learn something important on the much vexed question of elasticity."

"Weak hearts are by no means so common as is often supposed. Many a man who thinks he has got one is merely dyspeptic; many a woman owes her symptoms to tight lacing or insufficient feeding."

"A cool head and a steady nerve, with a smattering of electrical knowledge, are the prerequisites of a first-class lineman. Unless a man be apt to judge and quick to remember, he will be liable to serious blunders in his manipulations of the wires. Where there are dozens of them attached to one pole, he should be able readily to distinguish each from the others and have no doubt as to whether it be quick or dead. Contact with the earth through means of a conductor should be shunned like death itself. So long, however, as the lineman sits astride the cross-piece of a wooden telegraph pole and confines his attention solely to a wire that has no communication with the earth, he is as safe as if he were in his mother's arms."

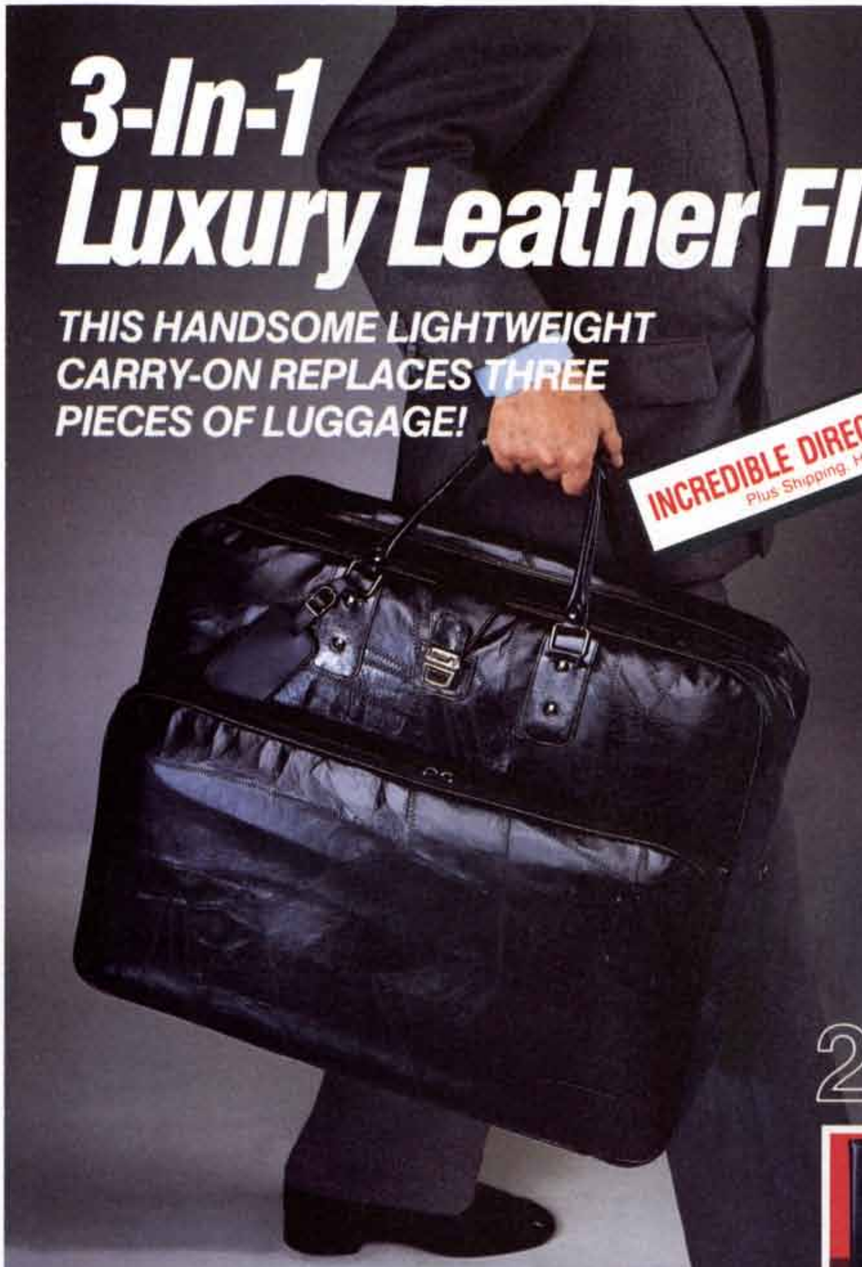
"Professor Fresenius, of Wiesbaden, after a long series of chemical analyses, declares that an egg contains as much nourishment as a pound and an ounce of cherries, a pound and a quarter of grapes, a pound and a half of russet apples, two pounds of gooseberries or four pounds of pears."

"The accompanying illustration represents the application of a system of electrical propulsion for common roads, by means of which traffic is designed to be carried on without employing a railroad track. It has been patented by Mr. Harvey D. Dibble, of Rapid City, Dakota Territory. Ordinarily the wagon will run in a line parallel with the overhead conductors, the trolley following any deviations from a straight line, and the driver not being required to use the steering lever except when it is needful to turn aside."

© 1989 SCIENTIFIC AMERICAN, INC

3-In-1 Luxury Leather Flight Bag.

THIS HANDSOME LIGHTWEIGHT CARRY-ON REPLACES THREE PIECES OF LUGGAGE!



1. ***IT'S A GARMENT BAG!*** *Holds 2 suits/skirts and ties.*

2. ***IT'S A SUITCASE!*** *Holds shirts/blouses, undergarments, shoes, socks & toiletries.*

3. ***IT'S A BUSINESS CASE TOO!*** *Holds files, writing supplies, pens, calculator and miscellaneous items.*

Introducing Europe's premier piece of luggage, the 3-in-1 Luxury Leather Flight Bag. This unique, practical carry-on is ingeniously designed to meet all of your travel needs. For an overnight trip, a five day conference or a vacation, 3-in-1 will perform for you like no other bag in the world! The 3-in-1 Luxury Leather Flight Bag is lightweight and supple yet remarkably durable. It is crafted of genuine cowhide leather sections, meticulously matched by hand and sewn together. This distinctive luggage is designed to be efficient and spacious. It has seven compartments and is made with many features not found in luggage three times its affordable price. It has everything you need for your travels. *Act today!*

30 DAY FREE TRIAL

YOUR GUARANTEE OF SATISFACTION!

Your complete satisfaction is our number one priority. That's why we invite you to inspect your 3-in-1 Luxury Leather Flight Bag for a *full* 30 days without risk or obligation. After 30 days, you can return the 3-in-1 Luggage (in its original condition) for a full and immediate refund. *No questions asked!*

Only $99

Plus Shipping, Handling & Insurance — $6

Available In Jet Black and Rich Chestnut Brown.

ORDER TOLL-FREE

1-800-558-1500 Ext. 13

VISA MasterCard

or send check or money order to:

JET SET™

1601 W. Glenlake Avenue, Dept. 13
Itasca, IL 60143

© 1989 SCIENTIFIC AMERICAN, INC

SCIENCE AND THE CITIZEN

Science Gains a Voice

A new appointment raises hopes for better funding of research

During the final years of Ronald Reagan's presidency, scientific advice in the White House shrank to near-invisibility. D. Allan Bromley of Yale University, an eminent nuclear physicist who was nominated as director of the Office of Science and Technology Policy (OSTP) three months after the Bush Administration took office, may change all that.

Bromley's unimpeachable scientific credentials, coupled with his reputation for outspokenness and his long-standing championship of funding for research, have renewed optimism that science will be clearly heard in the Bush White House. Besides being director of the OSTP, Bromley will also have the title of Assistant to the President for Science and Technology, a new position that gives him access to the Oval Office and participation in policy-making councils.

Bromley enters a game that has already begun. An array of mega-projects, including the space station, the Superconducting Supercollider and the human genome project, are competing for federal dollars with politically sacred nonscience programs on Capitol Hill; tentative budget compromises suggest science requests may not be fully met. Nevertheless, President Bush has renewed Reagan's long-stalled pledge to double funding for the National Science Foundation by 1993.

Bromley seems to feel more than comfortable with this expansionist agenda. He supports—as "an appropriate general goal"—a recent call by Frank Press, president of the National Academy of Sciences, for a doubling of the current $10.5-billion total federal investment in basic research over the next five years. Bromley too sees the funding of research as an investment: he cites an unpublished study by Edwin Mansfield of the University of Pennsylvania that estimates the social return on academic research worldwide in the past 15 years at some 28 percent per annum. Mansfield's preliminary data are "the first that give us a firm hold on this question," Bromley asserts.

He suggests he will want to scrutinize the balance between small-scale research and the big-ticket items that attract media attention. In particular, he questions whether some planned big science or technology projects should be all-American; perhaps, he suggests, efforts should be made to secure foreign participation before the construction stage.

Bromley's experience has prepared him for Washington. He has served on many government panels and was a vice-chairman of the committee that in 1986 first proposed doubling the budget of the National Science Foundation. He has earned a reputation for being forthright, even somewhat abrasive, and is said to have a deep regard for scientific discipline and disdain for hype. "In terms of preparation, he gets a very, very high mark," says William D. Carey, former executive officer of the American Association for the Advancement of Science. "I don't know that he would get many bouquets for congeniality, but his other assets are really central."

Certainly Bromley is willing to speak out: in 1980 he made headlines by calling attention to declining scientific literacy in the U.S., and he remains emphatic about the need to improve science education. At the precollege level, he says, it "can only be classed as scandalous"; at the college level there are "peaks of excellence and valleys of mediocrity that defy description," and it is only at the postgraduate level that the U.S. is a world leader.

Bromley thinks his most important job will be to give President Bush access to sound advice on any issue that involves science or technology and to "sharpen the options available to the President." Indeed, one of Bromley's first tasks will be to suggest names for the President's Council of Science and Technology Advisers, a new body that is intended to play a more active role than did its predecessor, the White House Science Council.

To be personally effective, Bromley suggests, he must earn the trust of President Bush and his advisers and thus come to be perceived as a member of the White House inner circle. With an eye to that goal, Bromley says he will leave any personal agenda in New Haven. Perhaps for the same reason, he declines to be drawn out concerning many topical science-and-society questions. He was, however, on the executive board of the International Council of Scientific Unions when the idea of a vast study of earth systems was first put forth, and so he is likely to be well informed on issues of global change. With that subject apparently growing more politically charged, the exposure could be useful. *—Tim Beardsley*

Cosmic Collisions

Computer simulations suggest how elliptical galaxies formed

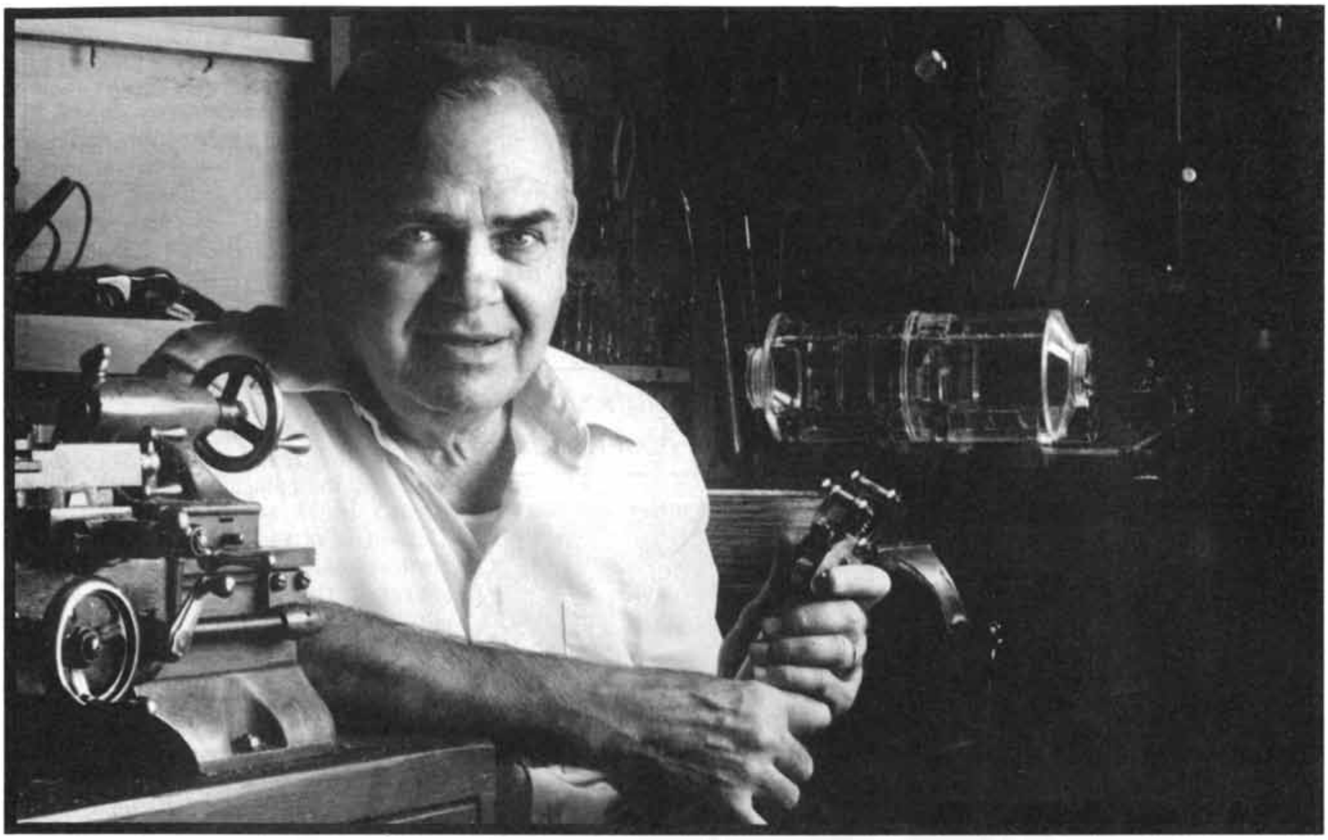
By ramming ordinary galaxies into each other in the deep space of a supercomputer, two astrophysicists have shown how a class of galaxies known as ellipticals might have formed.

More than two thirds of all galaxies, including spiral ones such as the Milky Way, are roughly disk-shaped. Elliptical galaxies, in contrast, are more full-figured, ranging in shape from ellipsoids to spheres. They tend to be larger than disk galaxies, and they often have "active nuclei": central regions spewing out prodigious amounts of radiation.

Not long ago many investigators thought that elliptical galaxies represented relics of an early phase of galaxy formation and that disk galaxies formed later. More recently some astrophysicists have advanced an alternative idea: that disk galaxies formed first and then occasionally collided to form elliptical galaxies. Computer calculations made by Joshua E. Barnes and Lars Hernquist of the Institute for Advanced Study in Princeton, N.J., lend support to this newer theory.

Barnes laid the groundwork for the findings. He has spent several years developing a computer program that simulates how gravitational forces ex-

© 1989 SCIENTIFIC AMERICAN, INC



"It's not easy for an astronaut to build anything up there...what they really needed was a fastener that would help them put those space structures together quick and easy.

My job is solving these puzzles. This time I designed a self-aligning bolt for them. But I couldn't help thinking those gloves they wear could be trouble. So I went home and did some work on my own. Cut up a cookie sheet and found an old gear in the garage—used them to make some modifications on a cordless screwdriver. Turned it into a push-button wrench that'll let anybody—astronaut or robot—screw that bolt together in seconds.

If you keep on thinking, one interesting idea just naturally seems to lead right on to the next one." *—Earl Cooney, Space Station, Industrial Engineer Advisor*

MCDONNELL DOUGLAS

A company of leaders.

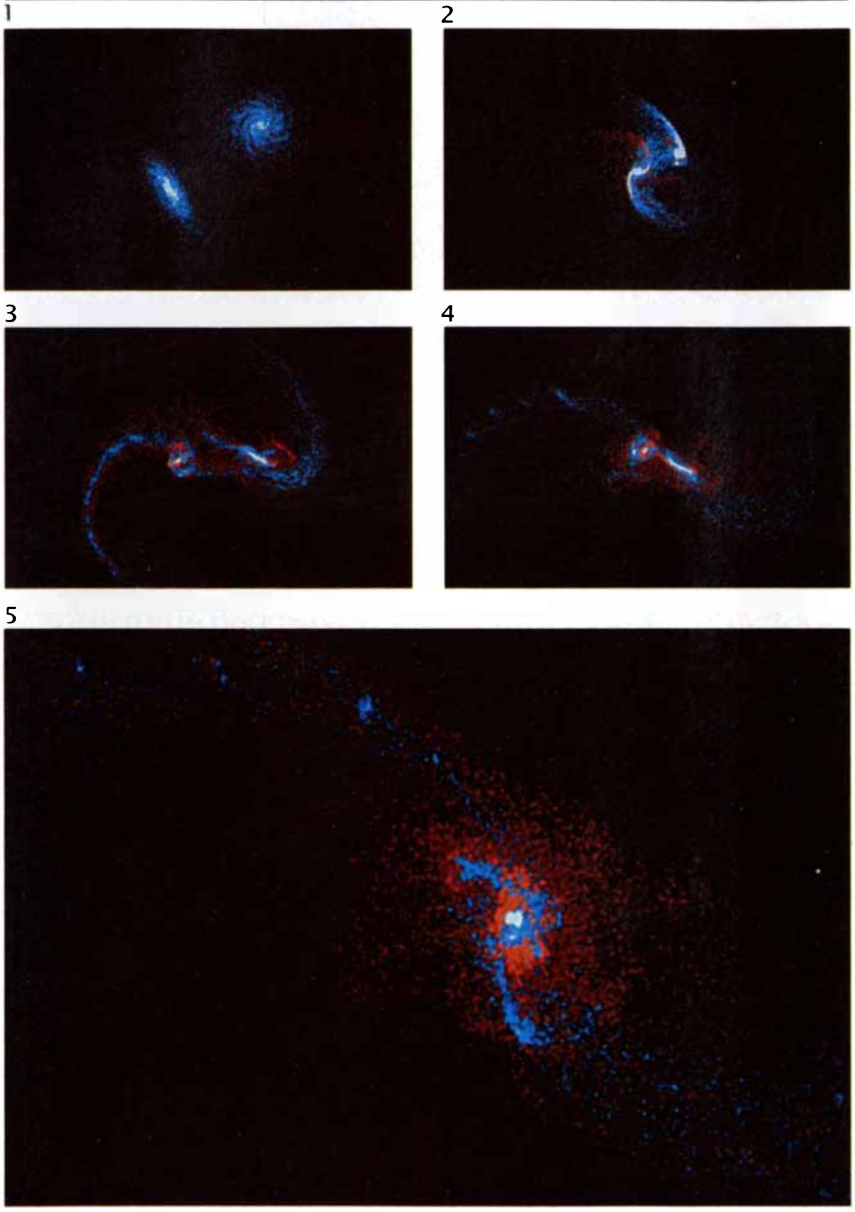
© 1989 SCIENTIFIC AMERICAN, INC

erted by stars and "dark matter"—thought to account for as much as 90 percent of the total mass of most galaxies—might affect interactions between two or more galaxies. Barnes recently concluded that if two galaxies pass within a few galactic diameters of each other (the Milky Way, to offer one example, is 100,000 light-years across) they might gravitationally "capture" each other and move inexorably toward a collision. A billion years or so after the initial collision, they would merge into a single object: an elliptical galaxy.

By folding interstellar gas into Barnes's model, Hernquist has arrived at a possible explanation of why some elliptical galaxies have hyperactive cores. Current theory holds that only huge amounts of matter plummeting into a massive black hole could generate active nuclei, which often outshine the rest of the galaxy. The question is, what supplies the matter to the black hole? Hernquist's calculations suggest that the collision of two disk galaxies would drive vast amounts of gas toward the center of the newly formed elliptical galaxy. This concentration of gas, Hernquist says, might spur the formation of a black hole and then provide it with a virtually endless reservoir of matter.

The work of Barnes and Hernquist has quelled many doubts raised initially by the collision theory of elliptical galaxies, according to François Schweizer of the Carnegie Institution in Washington, D.C. Indeed, Schweizer suggests that such simulations may provide a glimpse of our galaxy's future. The Milky Way, he points out, is approaching its nearest neighbor, the disk-shaped Andromeda galaxy, at a rate of 120 kilometers per second; a galactic crack-up could occur within five billion years. —*John Horgan*

Simulations of collisions between spiral galaxies generate elliptical galaxies that match observations



1 2 3 4 5

SPIRAL GALAXIES containing stars* (red), *gas* (blue) *and invisible "dark matter" collide in a simulation by Joshua E. Barnes and Lars Hernquist of the Institute for Advanced Study in Princeton, N.J. The merger's end product, which has a core of dense gas* (white) *and two long, curved "antennas," resembles some actual elliptical galaxies. The simulation was done at the Pittsburgh Supercomputing Center.

Moscow Subversive Chic

Dissent glues together the traditional chaos

Images are abundant: sidewalk artists displaying fluorescent nudes, a swing band belting out "Boogie-Woogie Bugle Boy"; militiamen carrying demonstrators into waiting buses, a scientist waving a copy of *Pravda* at the Presidium of the Academy of Sciences; teenagers in punk haircuts crossing themselves at an Easter service and above the altar a sign of a hundred light bulbs flashing on at midnight: "Christ has risen!"

Such images seem random and disconnected only if you have not been in Moscow recently. In Moscow, where anarchy traditionally undermines all order, anarchy itself has become the ordering principle; anything goes. Or does it? A walk through Ismailovsky Park confirms that *glasnost* has exposed a vacuum of experience. Thousands of identical paintings of churches, flowers, icons. The country is searching for an authentic voice that is more than merely a reaction to the establishment. It is a voice yet to be found; at the moment, to say what was previously forbidden—to protest—is to say enough. There is little surprise in that the most spontaneous performance on the Arbat is "Boogie-Woogie Bugle Boy"—an American

© 1989 SCIENTIFIC AMERICAN, INC

Mary Brodie won't feel the lump in her breast for another two years.

Like a lot of women, Mary Brodie understands the importance of regular breast self-examination. And because she's never felt a lump, she thinks everything is fine. It's the same conclusion a lot of women reach.

Unfortunately, it's wrong.

The tiny tumor that's forming in her breast is too small to feel. But with mammography, it's not too small to see.

And tomorrow Mary is getting her first mammogram. Thanks, in part, to a new x-ray film created by DuPont that makes it safer for women to start mammography early.

And for Mary, early detection means a two year head start on the rest of her life.

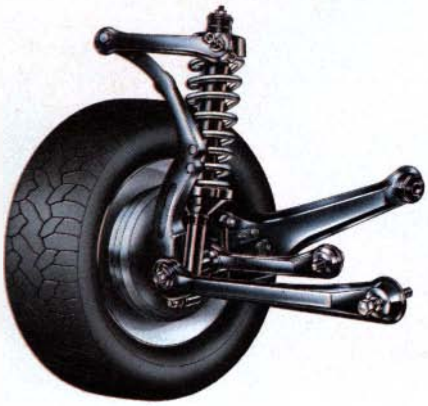
At DuPont, we make the things that make a difference.

Better things for better living. DUPONT

REG U S PAT & TM OFF

© 1989 SCIENTIFIC AMERICAN, INC

To make it
we had to tw



They said it couldn't be done.

That it would be impossible to create a suspension system that combined handling and ride without compromising either one. Then put it on a car that didn't cost a small fortune.

But the engineers at Honda ignored the naysayers. And found the answer: Arms.

The upper and lower arms keep all four tires virtually perpendicular to the road for outstanding handling and stability. Which frees up the springs and shock absorbers to do exactly what they were designed to do: absorb road shock. That way, you don't have to.

© 1988 American Honda Motor Co., Inc.

© 1989 SCIENTIFIC AMERICAN, INC

ride so well, ist a few arms.

You'll find this advanced double wishbone design on the Honda Accord. (And to your left). It was inspired by the double A-arms used on race cars. Honda was first to use them on all four wheels of a front-wheel-drive car. With a slight twist.

What all this means is that you will enjoy a comfortable, smooth ride as you put the responsive, 2.0 liter, fuel-injected 120 horsepower LXi engine through its paces.

And you'll appreciate the spacious interior, with standard air conditioning, full-logic AM/FM stereo cassette, power moonroof and driver's seat with adjustable lumbar support.

In the meantime, Honda engineers will keep looking for other innovative ideas.

It is a challenge they welcome with open arms.

HONDA

The Accord LXi

© 1989 SCIENTIFIC AMERICAN, INC

SCIENTIFIC AMERICAN
In Other Languages

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

サイエンス
Y950/copy Y10,440/year Y14,000/(abroad)
Editorial, subscription, advertising correspondence:
Nikkei Science, Inc.
No. 9-5, 1-Chome, Otemachi
Chiyoda-ku, Tokyo, Japan

INVESTIGACION Y CIENCIA
500 Ptas/copy 5500 Ptas/year 6200 Ptas (abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica S.A.,
Calabria, 235-239
08029 Barcelona, Spain

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

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

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

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

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

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

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

tune. Anything goes, but what goes is predictable.

I arrived in the Big Onion on April 8 for a working visit at *V Mire Nauki*, the Russian edition of *Scientific American*. It was my eighth trip in 10 years. The next day riots broke out in Tbilisi. Reports in *Pravda* and *Isvestia* were vague; rumors from Georgia were the first indications that the protesters had been killed by militiamen wielding spades. Despite *glasnost*, the Moscow rumor mill has, thankfully, not ground to a halt: "Civil war is approaching, you know."

On April 11 *Pravda* published the text of amendments to the criminal code that, among other things, made "discreditation" of leading officials of social organizations a punishable offense. The law bears an unsettling resemblance to the old Article 190[1], used to silence dissidents for "evident lying fabrications." The announcement of the amended criminal code and the riot reports did not immediately appear connected with the upcoming elections at the Academy of Sciences. Here I was mistaken; in Moscow everything is connected.

The Academy was then gearing up for the second-round elections to the Congress of People's Deputies, in which 12 seats left vacant after round one would be filled. The main question was whether Andrei D. Sakharov and the other liberal candidates removed by the Academy's elite before the first round would now be elected.

I was invited to the April 19 meeting in the Great Hall of Moscow University, where about 1,000 academicians, corresponding members of the Academy and representatives of the various institutes gathered to hear the platform of each of the 25 candidates and debate its merits. Form was beginning to intrude on chaos. The antiestablishment opened the meeting with protests on the university steps against the candidacy of Georgi A. Arbatov, director of the U.S.A. and Canada Institute, Moscow's best-known think tank on Western affairs. "Architect of stagnation," the placards read. "Stagnation" is standard *glasnost* for the Brezhnev era, a term encountered as often as paintings of churches.

Inside: academician after academician speaks out for Arbatov, openly questioning the motives of the protesters. Again the liberals: the image of a researcher waving a copy of *Pravda*'s article on "discreditation" at the Presidium is strong: "If we criticize members of the Academy's Presidium," he shouts, "are we in violation of the law?" The chairman's response reflects the code's and the country's state of ambiguity: "Criticize but don't discredit." Laughter is raucous—and uneasy with remembrance of 190[1]. The subject changes—but not the dialectic—and soon one speaker after another is taking the podium to denounce the news reports from Georgia as "lying and biased."

Any rules of order are lost on me; indeed, the election rules are lost on some academicians. Candidates with names at the end of the alphabet, such as Sakharov and Roald Z. Sagdeev, do not appear until late in the evening. Major points of Sakharov's platform include the prohibition of aboveground nuclear-power plants and the gradual abolition of the Soviet internal passport system. In agreement with many liberal thinkers, Sakharov suggests "the liquidation of the [economy's] administration-command system and its replacement with a pluralistic system, market controls and competition."

The elections are held the next day. The most votes (869 for; 232 against) go to Nikolai P. Shmelev, an economist who supports radical economic reform, including the "utmost development of market controls." Sakharov comes in sixth (806 to 295), Sagdeev eighth (739 to 362), Arbatov 11th (645 to 456). The list of round-two winners essentially duplicates the list of liberal candidates eliminated in round one.

To be liberal these days is all-important for the intelligentsia. At a meeting with a high-ranking cultural official, a secretary enters and asks if he is going to the Party meeting tonight. "Yes," he replies, then turns to me with a worried glance. "I'm not a Party Member you understand, but my position makes attendance mandatory."

So the elections, both within and beyond the Academy, were a clear victory for *glasnost* and protest—but for democracy? It is easy to estimate that the Academy's allocation of 20 seats out of the 2,250 in the new Congress makes each vote of the 70,000 scientists in the Academy worth about 20 votes of an ordinary citizen. Not exactly proportional representation, but still a great leap forward.

Moscow's directed anarchy was, of course, not confined to the Academy. On April 23 the Democratic Union, a radical organization that is allowed to exist but not to assemble, held a protest across from Pushkin Square in central Moscow. Some said it was a protest against the new criminal laws, others said against the events in Tbi-

© 1989 SCIENTIFIC AMERICAN, INC

lisi; the subject was not as important as the act. In any case, the militia had cordoned off the area and prevented 1,000 potential demonstrators from crossing the street to find out.

Five years ago we would all have been arrested. Indeed, I watched militiamen carry off a number of the demonstrators. What is the law? Even the militia seemed not to know. "Go home, take the metro, this is none of your business," was all they could come up with as the crowd shouted, "*Pozor! Pozor!* Disgrace! Disgrace!" "He who holds the megaphone is he who has *glasnost*," was my parting shot, drowned out I am sure.

April chaos is not restricted to art or politics, nor is protest. Religion is also an option. The ultimate act of subversion is to attend a Russian Easter service. At least the authorities seemed to think so in the past, barricading churches, disrupting the processions. This time, too, they made a show of it, and the militia surrounded Novodevichy Monastery as if they intended to seal it off. But it was a show: they checked neither passes nor passports: everyone got through. Inside, one's first impression was—anarchy. Foreigners everywhere, miniskirts, art-deco gold lamé (this lady might as well have been from Mars), teenagers holding hands and those—members of a motorcycle gang?

Was this a protest against authority or subversion without subversion, just as the militia's show was force without force? I don't know. It was largely curiosity or the search for a voice. Maybe we all just wanted to belong to something. But at midnight those hundred light bulbs flashed on. Where was I, Times Square? No, apparently Moscow. "Christ has risen!" "In truth he has risen!" —*Tony Rothman*

PHYSICAL SCIENCES

Stradivari's Secret (Redux)

Did baroque artisans "X-ray" wood with sunlight?

For hundreds of years musicologists have argued about why certain baroque instruments—violins made by Italy's Antonio Stradivari, for example, or harpsichords by the Ruckers family in Antwerp—produce such splendid sound. Some experts maintain the secret lies in the way the wood was cured; others claim it resides in the varnish.

WE'RE FIGHTING FOR YOUR LIFE

American Heart Association

ISSUED PRECISELY 20 YEARS AFTER THE APOLLO 11 LANDING

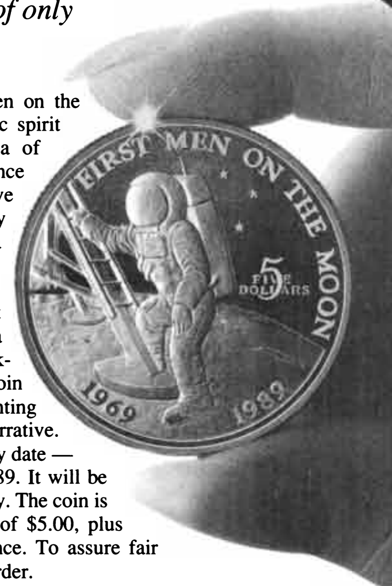
Official announcement of the

FIRST MEN ON THE MOON $5 COMMEMORATIVE COIN

An affirmation of American ingenuity and courage, available at the face value of only $5.00. Now you can hold a piece of history in your hand.

Twenty years ago, America landed the first men on the moon. And this year, the self-confident and heroic spirit that put America's flag on the moon's Sea of Tranquillity will be officially marked with the issuance of the First Men on the Moon $5 Commemorative Coin. This legal tender coin is being issued by the Republic of the Marshall Islands, where America has an important space tracking station.

The coin dramatically depicts the triumphant moment when American astronauts became the first men on the moon. About the same diameter as a Silver Dollar, the coin is even thicker — and painstakingly minted in a brilliant uncirculated finish. Each coin is protectively encased, after passing exacting minting standards, and is accompanied by an authoritative narrative.



Shown enlarged. Actual diameter: 1 ½"

This historic coin will be issued on the anniversary date — July 20, 1989 — and it will be minted only in 1989. It will be available only for a short time and in limited quantity. The coin is offered without any premium over its face value of $5.00, plus $1.00 per coin for shipping, handling, and insurance. To assure fair distribution, there is a strict limit of five coins per order.

Order promptly — for yourself or as a gift — from the Republic of the Marshall Islands Coin Fulfillment Center, One Unicover Center, Cheyenne, Wyoming 82008-1989. Phone TOLL-FREE 1-800-722-7121 from anywhere in the U.S. and Canada. All orders are subject to limitation and acceptance. *Satisfaction guaranteed.*

REPUBLIC OF THE MARSHALL ISLANDS COIN FULFILLMENT CENTER
One Unicover Center • Cheyenne, Wyoming 82008-1989

QTY Please accept my order as shown. Total price of $6.00 per coin includes shipping, handling, and insurance. Limit: 5.

Total amount of order: $____________.

❑ I enclose full payment by check or money order payable to Coin Fulfillment Center.
❑ Charge my order to
❑ American Express ❑ Visa
❑ MasterCard ❑ Diners Club

CARD NO EXP. DATE

SIGNATURE (All orders subject to acceptance)

MR./MRS./MISS/MS

ADDRESS

CITY

STATE ZIP

ORDER BY TOLL-FREE TELEPHONE, CALL

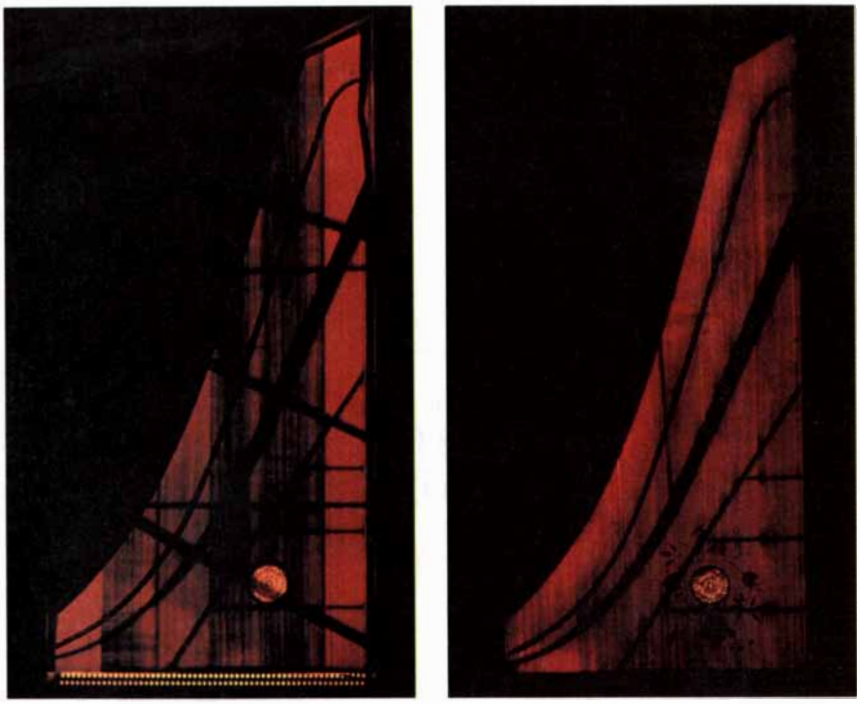
1-800-722-7121

from anywhere in the U.S. and Canada.
When calling, please mention this code: **BADY- FG01**

Your order will be acknowledged. Allow until September 20, 1989 for shipment.
©1989 RMI

© 1989 SCIENTIFIC AMERICAN, INC

Did "grain printing" help the masters find—and compensate for—variations in the density of wood?



HARPSICHORD SOUNDBOARDS made by Tillmann Steckner of London, Ontario, were transilluminated by electric light. Darker regions in the soundboard at the left, which has a uniform thickness, are more densely grained than surrounding regions. (The well-defined dark bands represent supporting struts.) Steckner thinned out densely grained regions in the soundboard at the right to give it a more uniform transparency—and, he claims, a better sound.

Here is another theory. Tillmann Steckner, a harpsichord maker in London, Ontario, thinks the baroque masters may have achieved their extraordinary success by using sunlight as a diagnostic tool. The notion first struck Steckner in 1974 when he casually held up a harpsichord soundboard to a window. The light shone through the sheet of spruce (the wood of choice for most keyboard and string instruments), revealing its internal structure. Although the wood's surface seemed to be uniformly grained, the backlighting showed that some regions were actually more densely grained—and darker—than others.

Later, while browsing through the scholarly literature, Steckner learned that the top plates of some baroque violins and the soundboards of some harpsichords have irregular variations in thickness. Thinking the variations were imperfections, restorers had occasionally planed down the wood to a more uniform thickness, but their efforts always ruined the instrument's sound.

Steckner wondered: had Stradivari and other masters used natural light to identify densely grained regions in their wood and then thinned them to compensate? Perhaps this technique improved the resonance of the wood by making the overall distribution of mass—rather than the thickness—more uniform. Steckner decided to test his theory—working not with sunlight but with electric light—on a harpsichord he was building for the St. Thomas Church in Leipzig, where Bach was once the music director.

Steckner claims the harpsichord, which he finished last fall, sounds better than any of his previous creations. Experts who have heard the instrument agree. "It has a quite amazing resonance," says Damjana Bratuz, a professor of music at the University of Western Ontario in London.

Steckner says he has learned of at least one other modern artisan—a German violin maker named Konrad Leonhardt—who has practiced what Steckner calls "grain printing," reportedly with success. But neither Leonhardt nor any other modern authority, to Steckner's knowledge, has proposed that the old masters employed the technique. "That's my contribution," says Steckner, who has described his research in a privately circulated manuscript.

Some experts doubt whether grain printing can produce an exceptional instrument. "The stiffness of the wood is what's important," observes Carleen M. Hutchins, who is an authority on the acoustics of violins, "and you can't equate stiffness with light transmission."

Steckner points out that the true test for his theory would be to determine whether relatively thin regions in the top plates of all Stradivari violins, for example, are also densely grained. Because spruce becomes opaque as it ages, such tests would require X rays. So far, Steckner adds, "No one has given me a Stradivari to take apart." —*J.H.*

Squaring the Circle

The sum of the parts equals the square or even the circle

Circles to square, and Cubes to double,/Would give a Man excessive Trouble.

—Matthew Prior, 1717

During the past two millennia an inconceivable number of compasses, straightedges and scissors have been bent and broken in the attempt to transform a circle into a square that covers the same area. Mathematicians gradually set these instruments aside as they proved ineffective. Yet 100 generations have not toiled in vain. The circle has been squared.

Miklós Laczkovich of Loránd Eötvös University in Budapest has demonstrated a method for partitioning a circle into a finite number of points, curves and twisted pieces that can be rearranged to form a square of the same area. "Many mathematicians who have worked on squaring the circle have reviewed the proof," says Stan Wagon of Smith College. "Laczkovich's method seems sound."

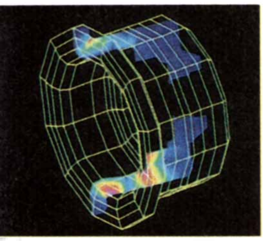
By applying a straightedge and compass, the ancient Greeks made the first attempt to construct a square from a circle. They failed, and so inevitably has every mathematician who has since tried this approach. A circle and square have an equal area only if the ratio between a side of the square and

© 1989 SCIENTIFIC AMERICAN, INC

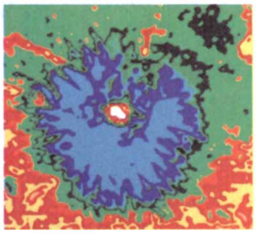
Genius should not go unrewarded.

CALL FOR PAPERS

THE 1989 IBM SUPERCOMPUTING COMPETITION



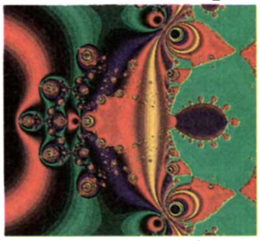
IBM invites you to submit your accomplishments. We seek to honor those individuals whose imagination has elevated the state of the art in analysis and modeling. Included are those who have contributed to creating new leading-edge uses of large-scale computers, innovative applications or algorithms and advances in numerically intensive computing.



The best papers on innovative uses of the IBM 3090 will be chosen in the following four divisions: physical sciences and mathematics; engineering; life and health sciences; social sciences, humanities and the fine arts. Each division will receive a first, second and third prize of $25,000, $15,000 and $10,000, respectively. In addition, a "PROCEEDINGS" of selected papers will be published.

Universities that provide substantial support to first prize papers will receive a $10,000 award as well.

Contestants must register an abstract of their paper, according to specifications, by October 2, 1989. Those who have properly registered their abstracts will then be invited to forward their papers by January 15, 1990.



To select the winners, IBM will retain a panel of independent authorities in each division. Winners will be announced by March 1, 1990.

Simply put, IBM would like to acknowledge those individuals who share the same interest we do...the unrelenting pursuit of knowledge.

For more information contact your IBM marketing representative, or call (914) 686-6318; in Canada, (416) 758-4136.

To obtain the general information package and an entry form write to:

Competition Administrator
IBM Corporation, Dept. 72/BNG
44 South Broadway
White Plains, NY 10601

Name ______

Affiliation ______

Address ______

City ______ State ______ Zip ______

The competition is open to any individual or team working or living in the U.S. or Canada, excluding employees of the IBM Corporation, IBM Canada LTD, and IBM related companies and their families. © 1989 IBM Corporation

IBM®

© 1989 SCIENTIFIC AMERICAN, INC

LIGHTNING DETECTION AND ADVANCE STORM WARNING SYSTEM

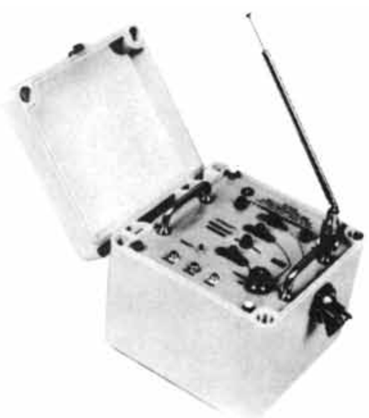
Reduce your liability and insurance claims, even death or injury. The SD-250B is effective in advance warning of unannounced "KILLER" storms.

This unique, state-of-the-art scientific instrument provides ample time to take evasive action against such unpredictable events. Solid state design consists of one IC, 31 transistors and 11 diodes.

The SD-250B satisfies the requirements of the U.S. Army Corps of Engineers, Bureau of Mines, OSHA and others interested in reducing casualties.

Lowest in cost, highest in quality and effectiveness, the SD-250B does what the weather bureau cannot.

The SD-250B is portable and is equipped with test calibrator, rechargable battery and battery charger.



ALARM RANGE:
10, 25, 50, 100 MILES (+/−20%)

SIZE:
8″ × 8″ × 6.5″

WEIGHT:
14 POUNDS

BATTERY:
RECHARGEABLE 6 VOLT GC-660
(CHARGER SUPPLIED WITH UNIT)

CASE:
WATERTIGHT MOLDED FIBERGLASS

OPTIONS:
REMOTE ANTENNA, STORM INTENSITY METER, CHART RECORDER, STROBE LIGHT, SIREN

SENSITIVITY:
1 (ONE) MICROVOLT W/70db LOAD

SDI INTERNATIONAL
LORTON, VIRGINIA 22079
(703) 550-9899
(800) 541-7250
FAX (703) 339-1269

a radius of the circle equals the square root of pi. Mathematicians proved in 1882 that two line segments cannot be constructed to have lengths in this ratio, and therefore this method cannot square the circle.

Undaunted by this mathematical impossibility, mathematicians abandoned the straightedge and compass in favor of the scissors. They tried to cut up a circle and rearrange the pieces to form a square. If you try to do so, you will soon discover that the round edges do not cooperate in the least. Indeed, after many attempts, you might convince yourself that the problem is impossible, which is exactly what Lester E. Dubins, Morris W. Hirsch and Jack Karush of the University of California at Berkeley proved 26 years ago.

Another version of the circle-squaring problem remained, however, one that Laczkovich ultimately solved: could any two objects of equal size be decomposed into a finite number of points, curves and other contorted pieces—those that cannot be cut using scissors?

Alfred Tarski, who first posed this problem, and Stefan Banach solved the three-dimensional case in 1924. They divided a ball into convoluted and perforated pieces that could be rearranged into a cube of the same volume. Paradoxically, a sphere can also be partitioned into pieces that can be recombined into virtually any shape of any volume [see "Computer Recreations," by A. K. Dewdney; SCIENTIFIC AMERICAN, April]. Remarkably, the two-dimensional version of this problem turned out to be more difficult.

Laczkovich seems finally to have unraveled the answer to it. In a 39-page proof he describes the unusual pieces from which circles and squares are made. In addition to the circle, his method can also square the ellipse or any other plane figure whose boundary can be represented by a smooth loop of untwisted string. He demonstrates that any of these plane figures can be decomposed into pieces that one can form into a square merely by moving them without rotation.

Some of the pieces resemble those of an ordinary jigsaw puzzle, but many are collections of single points or curved segments. You would not, however, want to assemble them yourself unless you have sufficient spacetime. Laczkovich estimates that his method of squaring the circle produces about 10^{50} pieces. If each of the pieces averaged one square centimeter in size, then a circle composed of the pieces would cover about a 10th of our galaxy. *—Russell Ruthen*

Jukeboxes for Scientists

NASA seeks ways to cope with an upcoming deluge of data

After years of yearning for new data, U.S. space scientists may soon have more than they know what to do with. Consider these figures:

During the past few decades the National Space Science Data Center—which is the National Aeronautics and Space Administration's main archive for data gathered by scientific spacecraft, from *Pioneer* to the *International Ultraviolet Explorer*—has accumulated some 6,000,000,000,000, or six trillion, bytes of data (one byte equals eight bits, or binary units). That is roughly double the amount of information contained in all of the Library of Congress's 19 million books.

In late April the space shuttle sent *Magellan* on its way to a rendezvous with Venus; by the time the spacecraft finishes mapping that planet's baked topography some three years hence, it will have added three trillion bytes to NASA's current pool of data, raising it by 50 percent.

Magellan's output will be a mere trickle compared with that of the *Hubble Space Telescope,* scheduled to be launched next year. The telescope should generate several trillion bytes every year, thereby increasing NASA's data base more than fivefold over its 15-year lifetime.

The flood will turn into a tidal wave when, or if, NASA deploys the *Earth Observing System* (EOS) by the mid-1990's. The battery of sensors planned for EOS would generate a trillion bytes' worth of data every few days. Of course, by then the *Cosmic Background Explorer,* the Jupiter probe *Galileo,* the *Gamma Ray Observatory* and a host of other missions should also be pouring data back to the earth.

Are NASA's preparations for the deluge adequate? Measured in terms of funding, not really, acknowledges James L. Green, director of the national data center. NASA has proposed increasing the center's annual budget of \$3.2 million by \$1.5 million next year. "That's just not enough" to do all that should be done, Green says. Given the tremendous constraints on NASA's budget, he adds, data management will probably always be short-shrifted. "NASA's primary responsibility is to

© 1989 SCIENTIFIC AMERICAN, INC

build the instruments and then get the damn things into orbit," he explains. Unless that happens, he points out, there will not be any data to manage.

NASA is nonetheless trying to upgrade its data-handling resources. The agency is extending its electronic links to research institutions so that more data can travel more quickly to more places. Researchers are devising software programs that can automatically label raw, incoming data so that the information can be easily retrieved later. Such a program, Green explains, might allow a scientist to call up spectroscopic data gathered by EOS only on cloudy days in the tropics, for example, without searching the entire data base.

At the national data center, plans call for augmenting the current storage system of computer tapes and disks with "optical-disk jukeboxes." Each jukebox would have a 100-billion-byte capacity and would serve up data on command to researchers who come to the center or who log in on computer networks.

In pursuing these ideas, Green says, NASA officials are consulting closely with the space-science community and with other outfits experienced in handling electronic data. Conveniently, he notes, the national data center is housed in the Goddard Space Flight Center in Greenbelt, Md., "right up the road" from the National Security Agency in Fort Meade, perhaps the world's premier intelligence-gathering organization. *—J.H.*

BIOLOGICAL SCIENCES

Molecular Archaeology

DNA from a 7,000-year-old brain opens new vistas in prehistory

Human evolution is a molecular process, and it would seem natural for the study of that process to be focused largely on molecules. Yet until recently archaeologists have been frustrated by the fact that only present-day human DNA sequences were available for study; those from the past had to be inferred. A recent technique called the polymerase chain reaction, or PCR, however, has made it possible to obtain nucleotide sequences from ancient DNA and compare them directly with modern ones. Among the first fruits of this method are DNA sequences from a 7,000-year-old brain that shed new

FREE SCIENCE BOOKS CATALOG

Send for your free catalog of new and recent publications from W. H. Freeman and Company, the book publishing arm of *Scientific American* for more than 20 years. Discounts and bonus books offered.

☐ Yes, I'd like a free copy of your latest science book catalog.

☐ Please send literature about the Scientific American Library, too.

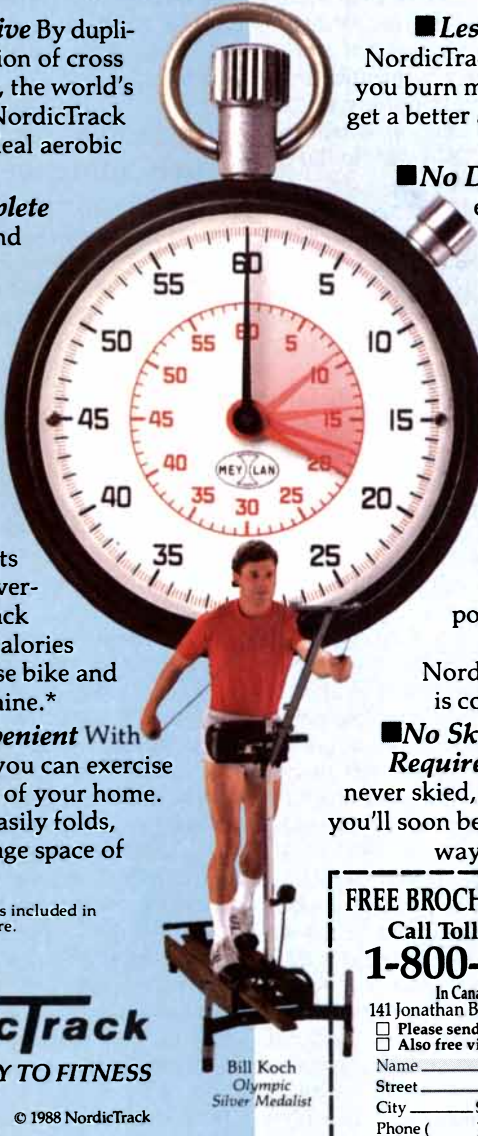
Name______

Address______

City, State, Zip______

Mail to: Gail Harleston
W. H. Freeman and Company, 41 Madison Avenue, New York, NY 10010

Exercise More with Less



■ ***More Effective*** By duplicating the motion of cross country skiing, the world's best exercise, NordicTrack provides the ideal aerobic workout.

■ ***More Complete*** Unlike bikes and other sitdown exercisers, NordicTrack exercises all the body's major muscles for a total body workout.

■ ***More Calories Burned*** In tests at a major university, NordicTrack burned more calories than an exercise bike and a rowing machine.*

■ ***More Convenient*** With NordicTrack, you can exercise in the comfort of your home. NordicTrack easily folds, requiring storage space of only 17" x 23".

*Scientific test results included in NordicTrack brochure.

■ ***Less Time*** Because NordicTrack is so efficient, you burn more calories and get a better aerobic workout in less time.

■ ***No Dieting*** No other exercise machine burns more calories than NordicTrack . . . So you can lose weight faster without dieting.

■ ***No Impact*** Running and some aerobic workouts can cause painful and potentially harmful jarring. A NordicTrack workout is completely jarless.

■ ***No Skiing Experience Required*** Even if you've never skied, in a few minutes you'll soon be "tracking" your way to better health.

NordicTrack
THE BEST WAY TO FITNESS
A CML COMPANY © 1988 NordicTrack

FREE BROCHURE AND VIDEO
Call Toll Free Or Write:
1-800-328-5888
In Canada 1-800-433-9582
141 Jonathan Blvd. N., Chaska, MN 55318
☐ Please send free brochure
☐ Also free video tape ☐ VHS ☐ BETA
Name______
Street______
City____ State____ Zip____
Phone ( )______ 320G9

© 1989 SCIENTIFIC AMERICAN, INC

light on the peopling of the Americas.

The power of PCR stems from the fact that it enables a biochemist to amplify, or copy, a particular stretch of DNA many times. Once the DNA has been amplified, the information it carries in the form of a sequence of nucleotide bases can be determined by well-established methods. As a result, DNA sequences—such as those from rare or damaged DNA—that were previously unobtainable can now be retrieved. PCR has had a revolutionary effect on a variety of fields, of which archaeology is only the latest.

The problem PCR overcomes in archaeology is that conventional methods of cloning (replicating) genes do not work well on ancient DNA. It has been known for some time that tissues of humans and other species are preserved in museum specimens, mummies and animals frozen in ice. But in all these tissues degradative processes have broken the DNA into small pieces and damaged it in other ways that render conventional sequencing methods impotent. In contrast, PCR can operate on small fragments of intact DNA within larger, damaged molecules.

Now Svante Pääbo of the University of California at Berkeley has exploited these properties to obtain DNA sequences from human remains found last year in Little Salt Spring, a bog in southwestern Florida. Radiocarbon dates show that the remains date from about 7,000 years ago, a time that falls in the so-called Middle Archaic period of Native American prehistory. Remarkably, because of the anaerobic and neutral conditions in the bog, the entire brain was preserved, albeit in a somewhat shrunken form.

The DNA Pääbo worked with comes from the mitochondria: organelles that supply a cell with its fuel. All of an organism's mitochondria are derived from the egg (rather than the sperm); hence they carry only maternal genes. Because mitochondrial genes mutate rapidly, they are valuable for tracing events in human prehistory, which occupies a brief span in evolutionary terms. (Pääbo's colleague Allan C. Wilson has analyzed modern mitochondrial DNA sequences and concluded that all modern human beings are descended from one woman who lived in Africa some 200,000 years ago.)

The DNA sequences from Little Salt Spring may ultimately help increase anthropologists' estimates of how many people initially immigrated to the Americas. It is widely thought that the original migration from Asia to Alaska was quite small, a conclusion buttressed by the fact that among modern Native Americans only three mitochondrial DNA lineages were known. Pääbo and his co-workers found that the brain from the bog contained a previously unknown mitochondrial sequence, which implies the existence of a fourth lineage that died out some time after the initial immigration.

Even beyond the specific findings from Little Salt Spring, PCR may open up new avenues of research in archaeology. Pääbo and his colleagues have used the technique to get DNA sequences from the body of an Egyptian priest (4,000 years old) and from a mammoth frozen in the Siberian ice for 40,000 years; other projects are already under way. As Pääbo notes in a recent paper in the *Journal of Biological Chemistry:* "Advances in molecular biological techniques have enabled us to retrieve... ancient DNA molecules and thus to catch evolution red-handed." *—John Benditt*

Alien Influence

Parasitic intruders manipulate host behavior

The notion that evolution shapes animal behavior has become familiar, but the idea that evolution can lead to the direct manipulation by a parasite of its host's behavior is something else again. Yet such control seems to have been demonstrated in an insect parasite by Jacques Brodeur and Jeremy N. McNeil of Laval University in Quebec. They have found that the larvae of a parasitic wasp change the behavior of their host species, the potato aphid, in different ways depending on the parasite's particular interests.

The female of the wasp *Aphidus nigripes* injects her eggs into the potato aphid; when the larvae hatch, they slowly consume the unfortunate aphid from the inside and so kill it. When the larvae have completed their development, they emerge from the mummified remains of the aphid. If winter is coming on, however, the parasite (properly termed a parasitoid, because it kills its host) may stay within the remains of the aphid, entering a prolonged period of suspended animation called diapause, until conditions are more favorable in the spring.

In *Science,* Brodeur and McNeil report that unparasitized aphids generally stay under leaves, where they prefer to feed, but that parasitized ones move to the leaves' upper surfaces before dying. The workers began their study by adjusting the conditions in laboratory colonies of parasitized aphids so that roughly half of the parasites would enter diapause.

Brodeur and McNeil then investigated whether the effect of a diapausing wasp larva on a host differs from the effect of a nondiapausing larva. The answer was a clear "yes": aphids parasitized by diapausing wasp larvae tend to leave the plant and hide themselves in concealed sites before dying, whereas aphids carrying nondiapausing larvae stay on the plant.

Many parasites are known to affect host behavior; the new twist is that *A. nigripes* is selective, making its host hide when hiding is adaptive for the parasite and not otherwise. Brodeur and McNeil suggest that diapausing larvae induce their hosts to lodge in sheltered places before dying so that the parasite will be protected against temperature changes and the attentions of predators during the winter.

They demonstrated the second point with a hyperparasitoid wasp species: one that places its larvae inside already parasitized, mummified aphids, where they eat the original parasite. Brodeur and McNeil showed that aphids that had been parasitized by diapausing wasp larvae—and were therefore concealed—were less frequently attacked by hyperparasitoids. Potato aphids not only are killed by *A. nigripes* but also are made to protect their executioners. *—T.M.B.*

Ripe for a Change

A new treatment could mean the end of "pink bullet" tomatoes

Anyone who has tasted a tomato ripened on the vine knows that the flavor of most supermarket tomatoes does not even come close. The reason is that most tomatoes are picked while they are still completely green, in order to give shippers time to get them to market. Such tomatoes are ripened artificially with the gas ethylene, but they never develop the full flavor and texture of tomatoes picked after they have begun to ripen. Peter J. Davies of Cornell University calls these artificially ripened tomatoes "pink bullets."

He and his colleagues think they have found a way to make pink bullets part of gastronomical history: a treatment that, they say, doubles the time

© 1989 SCIENTIFIC AMERICAN, INC

The ultimate in sound...
The ultimate in savings...

GET 6 COMPACT DISCS, CASSETTES OR RECORDS for the price of 1

...with nothing more to buy ever!

Bon Jovi: New Jersey 100516

100586 200596 100711

100715. R.E.M.: Green Orange Crush, Pop Song 89, etc. (Warner Bros.)

100602. Elton John: Reg Strikes Back • Elton's 22nd gold album! (MCA)

264134. D.J. Jazzy Jeff & The Fresh Prince: He's The D.J., I'm The Rapper (Jive)

200478. Metallica: And Justice For All • #1 Speed metal band! (Elektra)

223559. The Beach Boys: Endless Summer • 21 timeless hits! (Capitol)

100532. Diane Schuur: Talkin' 'Bout You • Cry Me A River, etc. (GRP)

100603 115436 100707

105392. Pops In Space John Williams & The Boston Pops. Music from Star Wars, The Empire Strikes Back, more. (Philips *DIGITAL*)

153582. Tracy Chapman Fast Car, Talkin' 'Bout A Revolution, etc. (Elektra)

164165. Bobby McFerrin: Simple Pleasures • Don't Worry Be Happy, etc. (EMI)

244006. Simon & Garfunkel: The Concert In Central Park • All-time classics! (Warner Bros.)

125179. Tchaikovsky, 1812 Overture; Romeo And Juliet; Nutcracker Suite Chicago Symph. Orch./ Solti. (London *DIGITAL*)

100459. Cocktail (Original Soundtrack) • (Elektra)

100604. Heifetz: The Decca Masters, Vol. 1 Golliwog's Cakewalk, Clair de lune, many more. (MCA)

100035. Robert Palmer: Heavy Nova • Simply Irresistible, More Than Ever, etc. (EMI)

123385. The Best Of Eric Clapton: Time Pieces (Polydor)

100579. K. T. Oslin: This Woman • Hold Me, Money, title song, more. (RCA)

100470. Vangelis: Direct New Age Meditations, The Motion Of Stars, The Will Of The Wind, etc. (Arista)

153983. Charlie Parker: Compact Jazz • Now's The Time, Night And Day. (Verve)

154135. The Best Of Steely Dan: Decade 14 hits. (MCA)

104871. Supertramp: Classics (14 Greatest Hits) • The Logical Song, Give A Little Bit, more. (A&M)

144578. The Judds': Greatest Hits • (RCA)

115356. Pinnock: Vivaldi, The 4 Seasons • Simon Standage, violin; etc. (Archiv *DIGITAL*)

114780. Cinderella: Long Cold Winter • Gypsy Road, Don't Know What You Got, more. (Mercury)

134347. Huey Lewis: Small World • (Chrysalis)

173406. Jazz CD Sampler Over 67 minutes of jazz. (Polygram)

100467. Beethoven, Symphony No. 9 (Choral) London Classical Players/ Norrington. (Angel *DIGITAL*)

123721. Jimmy Page: Outrider • Led Zeppelin guitarist's solo flight! (Geffen)

134321. Led Zeppelin: Houses Of The Holy (Atlantic)

153606. INXS: Kick • Need You Tonight, Devil Inside, etc. (Atlantic)

100517. Phil Collins: Buster/Soundtrack • Groovy Kind of Love, Two Hearts, etc. (Atlantic)

134420. John Cougar Mellencamp: The Lonesome Jubilee • Paper In Fire, more. (Mercury)

100008. Randy Travis: Old 8x10 • Honky Tonk Moon, more. (Warner Bros.)

100927 115457 100713

182522 115311 170348

270106. An Evening With Louis Armstrong • (GNP Crescendo)

262889. Virgil Fox: The Digital Fox • Organ music by Bach, others. (Bainbridge *DIGITAL*)

209468. Perlman: Brahms, Violin Sonatas (Angel *DIGITAL*)

120768. 20 Great Love Songs Of The 50s & 60s, Vol. 1 • (Laurie)

153740. Genesis: Invisible Touch • (Atlantic)

163579. Andrés Segovia Plays Rodrigo, Ponce & Torroba • Fantasia para un Gentilhombre, Concierto del Sur, Castles Of Spain. (MCA)

100679. Steve Earle: Copperhead Road • (UNI)

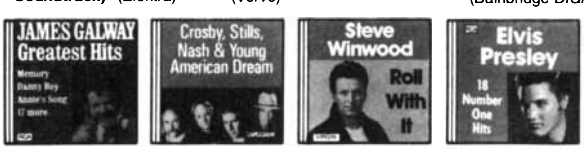
134267. Marriner: Mozart, Overtures • Academy of St. Martin. (Angel *DIGITAL*)

125360. By Request...The Best Of John Williams & The Boston Pops • Olympic Fanfare, Liberty Fanfare, more. (Philips *DIGITAL*)

134627. Classic Old & Gold, Vol. 1 • 20 hits! (Laurie)

104857. Benny Goodman: Sing, Sing, Sing • (RCA)

115306. Pinnock: Handel, Water Music • The English Concert. "A winner."—Ovation (Archiv *DIGITAL*)



173233 100714 154633 172190

Start With 4 Compact Discs, Cassettes or Records For 1¢! Yes, pick any 4 hit albums for only 1¢. You agree to buy only 1 more at regular Club prices (usually $14.98–$15.98 for compact discs, $8.98–$9.98 for cassettes and records)...and take up to *one full year* to do it. Then you can take another album free. That's 6 hits for the price of one and there's *nothing more to buy ...ever!* Shipping and handling added to each shipment.

How The Club Operates You select from hundreds of exciting hits described in the Club's magazine and mailed to you approximately every 3 weeks (19 times a year). Each issue highlights a Main Selection in your preferred music category, plus alternate selections. If you'd like the Main Selection, do nothing. It will be sent to you automatically. If you'd prefer an alternate selection, or none at all, just return the card enclosed with each issue of your magazine by the date specified on the card. You will have at least 10 days to decide, or you may return your Main Selection at our expense for full credit. Cancel your membership at any time after completing your membership agreement, simply by writing to us.

Free 10-Day Trial Listen to your 4 introductory selections for a full 10 days. If not satisfied, return them with no further obligation. You send no money now, so complete the coupon and mail it today.

MS302A BMG Music Service, 6550 E. 30th St., Indianapolis, IN 46219-1194.
TRADEMARKS USED IN THE ADVT ARE PROPERTY OF VARIOUS TRADEMARK OWNERS

BMG MUSIC SERVICE

Return this coupon...today!

❶ TAPE PENNY HERE

Mail to: BMG Music Service
P.O. Box 91001, Indianapolis, IN 46291

I enclose 1¢. Please accept my membership in the BMG Music Service and send me the 4 hits I've indicated here under the terms in this ad. I agree to buy just 1 more hit at regular Club prices, and take up to a year to do so—after which I can choose a FREE bonus hit! (Shipping/handling added to each shipment.)

❷ **SEND MY SELECTIONS ON** (check **one** only):
☐ **COMPACT DISCS*** ☐ **CASSETTES** ☐ **RECORDS**

❸ **I am most interested in the following type of music–but I am always free to choose from every category** (check **one** only): 1 ☐ **EASY LISTENING** 2 ☐ **COUNTRY** 3 ☐ **HARD ROCK** 4 ☐ **POP/SOFT ROCK** 5 ☐ **CLASSICAL**

❹ **RUSH ME THESE 4 SELECTIONS NOW** (indicate by number): ____ ____ ____ ____

❺ ☐ Mr. ☐ Mrs. ☐ Miss ________ First Name ________ Initial ________ Last Name ________ (PLEASE PRINT)

Address ________ Apt. ________

City ________ State ________ Zip ________

Telephone (________) ________ Area Code

U2W (AF)

Signature ________

❻ Have you bought anything else by mail in ☐ **last 6 months** ☐ **year** ☐ **never**

*Members who select compact discs will be serviced by the Compact Disc Club. Full membership details will follow, with the same 10-day, no-obligation privilege.

YGU2W (BU)

We reserve the right to request additional information or reject any application. Limited to new members, continental U.S.A. only. Current CD Club members not eligible for this offer. One membership per family. Local taxes, if any, will be added.

© 1989 SCIENTIFIC AMERICAN, INC

ripe tomatoes remain in top condition. Currently that time is four or five days. Davies says his technique extends the period to from 10 to 12 days. If the treatment proves economical, it should allow growers to pick the fruit after it begins to ripen and so eliminate the need for ethylene.

Davies and his colleague Martha A. Mutschler obtained their inspiration from a small Brazilian tomato called Alcobaça, which has a shelf life of several weeks. They and their collaborators found that Alcobaça contain nearly three times as much of a chemical called 1,4-butanediamine as most other types of tomatoes. The investigators then showed that by adding a very small quantity of the chemical to U.S.-grown strains they could double the tomatoes' shelf life.

The chemical is forced into the tomatoes by placing them in a dilute solution, lowering the ambient pressure and then suddenly releasing the vacuum. A small quantity of the solution enters the tomatoes through the stem scar.

If commercial-scale trials now under way prove successful, Davies points out, regulatory approval should be relatively straightforward because the amount of chemical added per fruit—a milligram—is similar to the amount present naturally. Moreover, other polyamines and even the amino acid methionine have the same effect. Although 1,4-butanediamine is a known plant hormone, its property of prolonging shelf life seems to be a separate effect, according to Davies. How it works is not yet understood, he admits; investigations are continuing. *—T.M.B.*

Sweet Success

A critical glucose transporter has been found

A discovery made virtually simultaneously by several groups has advanced understanding of a major question in metabolism research: what constitutes the full cascade of events leading from the binding of insulin by its receptor on fat and muscle cells to the cells' avid uptake of glucose from the blood? Such uptake is critical: it provides energy to a significant fraction of the body's cells, and it prevents glucose in the blood from accumulating to levels associated with diabetes.

The workers have isolated the gene for—and thus demonstrated the existence of—a protein, known as a glucose transporter, that appears to be crucial in the final steps of the cascade. In response to insulin's message, molecules of the transporter move en masse from the interior of the cell to the plasma membrane, where they usher glucose into the cell.

The search for the gene began last year, precipitated by work done by David E. James and Paul F. Pilch of the Boston University School of Medicine. They had been studying the activity of what was thought to be the only glucose transporter in fat and muscle cells: a protein initially identified in erythrocytes (red blood cells) and subsequently found in a wide range of cells. This protein had been shown to respond to insulin by migrating from interior pools known as microsomes to the plasma membrane. James and Pilch hoped an analysis of other proteins in the microsomes might provide clues to the migration mechanism and perhaps explain why too few of the transporter molecules appeared in the membrane to account readily for all the glucose taken up by insulin-stimulated cells.

To their surprise, one of the monoclonal antibodies they isolated to distinguish among microsomal proteins picked out a molecule that resembled the known transporter. It was about the same size, for example, and in fat cells exposed to insulin it traveled from microsomes to the plasma membrane; in fact, significantly more of the novel protein than of the erythrocyte-type transporter appeared in the membrane.

Definitive proof that James and Pilch had indeed identified a second, and potentially more important, glucose transporter awaited the isolation and characterization of its gene. At least five research groups have now independently done just that.

James, who is now at the Washington University Medical School, and his co-worker Mike Mueckler have cloned the rat gene and shown that its nucleotide sequence is related to, but clearly distinct from, the sequence of the erythrocyte-type transporter; so have Morris J. Birnbaum of the Harvard Medical School and, separately, Harvey F. Lodish and his colleagues at the Whitehead Institute for Biomedical Research. A similar sequence has also been identified in the mouse gene by M. Daniel Lane and his group at the Johns Hopkins University School of Medicine and the East Carolina University School of Medicine and in the human variety by Graeme I. Bell of the University of Chicago and his co-workers at several institutions. Birnbaum has also shown conclusively that the protein encoded by the cloned DNA does indeed transport glucose.

Workers are now hoping to identify the molecules that interact with the transporter to regulate its response to insulin. They are also examining the role of the transporter in non-insulin-dependent diabetes, the form that typically arises in adulthood. In particular, the possibility that an inborn defect in the transporter predisposes some people to diabetes is, as Birnbaum notes, too important to ignore. *—Ricki Rusting*

PROFILE

Ethnography as Art

"Complexifier" Clifford Geertz ponders anthropology's future

It once seemed so simple, not to say romantic: the intrepid ethnographer goes off to live with some exotic tribe (preferably isolated on a balmy South Sea island), records what he or she finds there and from those observations draws general conclusions about human behavior and social structure. Then doubts—moral and scientific—began cropping up. Does viewing other societies as "elemental" or "primitive" versions of so-called modern (that is, white, industrialized) society smack of imperialism and racism? Does the structure of any particular culture spring more from unique environmental influences than from some universal human traits? Does the investigator's own cultural indoctrination warp his or her observations? And finally, does the very act of writing distort "the truth"—whatever that is—even further?

Perhaps no anthropologist has posed these questions as pointedly and as persistently as Clifford Geertz. In *Works and Lives: The Anthropologist as Author,* the latest of more than a dozen books he has composed, Geertz sums up the sense of crisis that pervades his field: "Anthropologists have added to their 'Is it decent?' worry (Who are *we* to describe *them?*) an 'Is it possible?' one (Can Ethiopian love be sung in France?), with which they are even less well prepared to deal." Then, as he has for the past three decades, both by example and explicitly, Geertz points to a way out of the impasse. It is not to saddle anthropol-

© 1989 SCIENTIFIC AMERICAN, INC

CARRY WIZARD POWER IN YOUR POCKET. SLIDE IN A SOFTWARE CARD. OR PLUG INTO YOUR PC. TAKE CONTROL OF TOMORROW. TODAY.

The Wizard from Sharp lets you travel light and take all your personal and business data with you. The Wizard plans, organizes, reminds, informs, advises, calculates. An optional Organizer Link keeps you in two-way communication with IBM®-PC compatible or Macintosh® computers. Optional Software Cards adapt it for specialized jobs without specialized training.

Discover the competitive edge for the Information Age. Call 1-800-BE-SHARP.

SHARP INVENTS THE WIZARD

SHARP

FROM SHARP MINDS
COME SHARP PRODUCTS™

© 1989 Sharp Electronics Corp.

© 1989 SCIENTIFIC AMERICAN, INC

THE SCHOOL THAT COMES TO YOUR CHILD.

Kindergarten through 8th Grade.

High-quality, effective home study courses developed by certified teachers at outstanding Baltimore private school. Home is your classroom, you are the teacher. Success is easy with explicit, step-by-step instructions. Ideal for enrichment or for families on the move. All materials included. Start any time. Transfer to other schools. 350,000 student users in over 80 years. Non-profit. Equal opportunity. Fully accredited.
Write or call for free catalog.

CALVERT SCHOOL
Established 1897 (301) 243-6030

Dept. SCI79, Tuscany Rd., Baltimore, MD 21210

LOOKING FOR A PUBLISHER?

Did you know you can have your book published for a fee?

You are invited to send for a free illustrated, 40-page brochure which outlines the publishing services that have made Vantage Press America's leading subsidy publisher. This highly informative brochure will show how you can arrange to finance the prompt publication of your book.

To the Author in Search of a Publisher

Unpublished authors, especially, will find this booklet valuable. For your free copy, write to:
VANTAGE PRESS, Inc. Dept. F-103
516 W. 34th Street, New York, NY 10001

ogy with a specific ideological or political purpose or to ape the rigorous methodology of the "hard" sciences. Rather, Geertz says, the solution lies in accepting cultural anthropology as, for better or worse, a literary enterprise—"imaginative writing about real people in real places at real times." "Geertz is not the only spokesman for these ideas," says George E. Marcus of Rice University, "only the most original and brilliant."

This spring I interviewed Geertz on a drizzly day at the Institute for Advanced Study in Princeton, N.J., his home base since 1970. At 62, he is a big bear of a man, with shaggy gray hair and beard. His speech echoes his writing: all starts and stops, headlong assertions punctuated by mumbled asides. He is physically fidgety as well, pulling on his ear, pawing his cheek, slouching down in his chair and abruptly drawing himself up. Now and then, as he listens to me pose a question, he pulls the top of his sweater halfway up his face, like a bandit trying to conceal his identity.

His life story, too, seems at points intended to mystify rather than clarify. He grew up in Marin County, Calif., in "a sort of, kind of foster home," and his earliest ambition was "to get the hell out of there, we don't have to go into why." He succeeded by going to Antioch College in Ohio, where he majored in philosophy and fantasized about becoming a novelist. "I was the literary type," he recalls. "Of course my enemies would say I still am."

He became an anthropologist "quite accidentally." He intended to pursue philosophy after college, but then a professor at Antioch convinced him to look into the anthropology program at Harvard University. "I knew nothing about the field," Geertz recalls, but while he was in Cambridge a friend introduced him to Margaret Mead. The famous anthropologist spent an entire afternoon with Geertz, showing him notes from her fieldwork and regaling him with stories. He decided then and there to become an anthropologist.

His first fieldwork took him to Java in 1952, and since then he has spent dozens of months studying societies in Southeast Asia and in North Africa at firsthand. Indeed, Geertz points out with some fervor that although he is best known—especially outside the field of anthropology—for his theoretical musings, they account for only about 10 percent of his oeuvre; the bulk of his publications consists of "empirical" observations.

Geertz's most famous display of his brand of empiricism is "Deep Play: Notes on the Balinese Cockfight." The first sentence of the 1972 essay establishes his anything-but-straightforward style: "Early in April of 1958, my wife and I arrived, malarial and diffident, in a Balinese village we intended, as anthropologists, to study." (Geertz says his writing has been likened to both Marcel Proust and Henry James; he is flattered by the former comparison but thinks the latter is closer to the mark.) The opening section describes how the young couple gained the confidence of the normally aloof Balinese: the police raided a cockfight, and Geertz and his wife, who had been watching, fled along with the guilty villagers; impressed that the two Americans did not seek privileged treatment from the police, the villagers accepted them.

Having established his credentials as an insider, Geertz proceeds to depict and then to analyze the Balinese obsession for pitting roosters armed with razor-sharp spurs against each other. He eventually concludes that the bloody sport mirrors the Barlineses' fear of the dark forces underlying their superficially calm society. Like *King Lear* or *Crime and Punishment,* cockfighting "catches up these themes—death, masculinity, rage, pride, loss, beneficence, chance—and [orders] them into an encompassing structure."

One year later Geertz sought to justify his highly literary method in an essay titled "Thick Description: Toward an Interpretive Theory of Culture." An anthropologist, he maintained, can no more portray a society by "recording the facts" than a literary critic can get at the essence of a Shakespeare play by summarizing the plot. The anthropologist's calling, rather, is to "interpret" the culture, to extract meanings from it that are, ideally, as complex and richly imagined as the culture itself. This technique, which Geertz calls "thick description" (a term borrowed from the philosopher Gilbert Ryle), can never lead to a grand unified theory of human nature, but it can lead to a kind of provisional understanding—or, as Geertz puts it, in his characteristic deprecating fashion, to an increase in "the precision with which we [anthropologists] vex each other."

Geertz's ideas have found favor among anthropologists and other social scientists, as well as historians, philosophers and literary critics. But he vexes at least two types of anthro-

© 1989 SCIENTIFIC AMERICAN, INC

pologists. One includes those who think the field should be as quantitative and even predictive as physics. Geertz, who calls anthropology a soft or "modest" science, has little patience for this viewpoint. "Anthropology is responsive to evidence, and it theorizes," he explains. "But it has nothing like the status of the harder parts of the hard sciences, and I don't think it ever will."

Another, sometimes overlapping group of critics views anthropology as a tool to expose unjust social structures and identify superior (usually socialist) ones. These investigators accuse Geertz of treating societies as art objects and turning anthropology into a form of aesthetic appreciation, devoid of moral purpose. He rejects that assessment. "I don't consider myself a mere entertainer," he says. His work serves society, he contends, by providing "a more realistic, less Platonic view" of other cultures. He notes that in writing about Islam, for example, he has tried to show that it is "not an undifferentiated bloc of wild fanatics" and thereby to increase the potential for discourse.

Geertz recognizes as acutely as anyone the pitfalls inherent in a self-consciously literary approach to ethnography. "I always felt it could end in total failure," he says. Some younger anthropologists, he notes, have become so determined to expose all their potential biases, ideological or rhetorical, that their writings resemble confessionals, revealing far more about the author than about the putative subject. This trend, which Geertz dubs "I-witnessing," has produced some interesting works as well as some abysmal ones. But he insists that anthropologists should not let "epistemological hypochondria" drive them into a formal academicism, in which "anthropology will become only for anthropologists."

Noting that the "diffusion" of anthropology is one of its strengths, Geertz suggests that he would be horrified if all his colleagues suddenly decided to imitate him. "I don't believe in schools or system-building," he says. His goal has never been to perpetuate a specific methodology, he explains; it is rather to "set a tone or mood or agenda that people could react toward or against." When I tell him that another anthropologist has called him a "complexifier," he seems pleased. "I once wrote—not to quote myself—but I once wrote that any simplicity lies through complexity," he says, and smiles. *—John Horgan*

Who gives you over three decades worth of great scientific discoveries in one entertaining, fascinating book? Who else but...

FLANAGAN!

As editor of *Scientific American,* Dennis Flanagan spent thirty-five years bringing back reports from the frontiers of science. Now he shares them with fortunate readers in this "lively, immensely readable, wide-ranging overview" (*Newsday*).

"A pleasure... An utterly charming look at the present state of science by a 'spectator' who has a broader view and a deeper understanding than many of the practitioners do."
—ISAAC ASIMOV

"It succeeds beautifully. Mr. Flanagan's descriptions occasionally rise to the level of poetry."
—N.Y. Times Book Review

Now at your bookstore
VINTAGE BOOKS
A Random House Company

There's an epidemic with 27 million victims. And no visible symptoms.

It's an epidemic of people who can't read.

Believe it or not, 27 million Americans are functionally illiterate, about one adult in five. Forty-seven million more are able to read on only the most minimal level. Together, that's almost 75 million Americans...one third of our entire population.

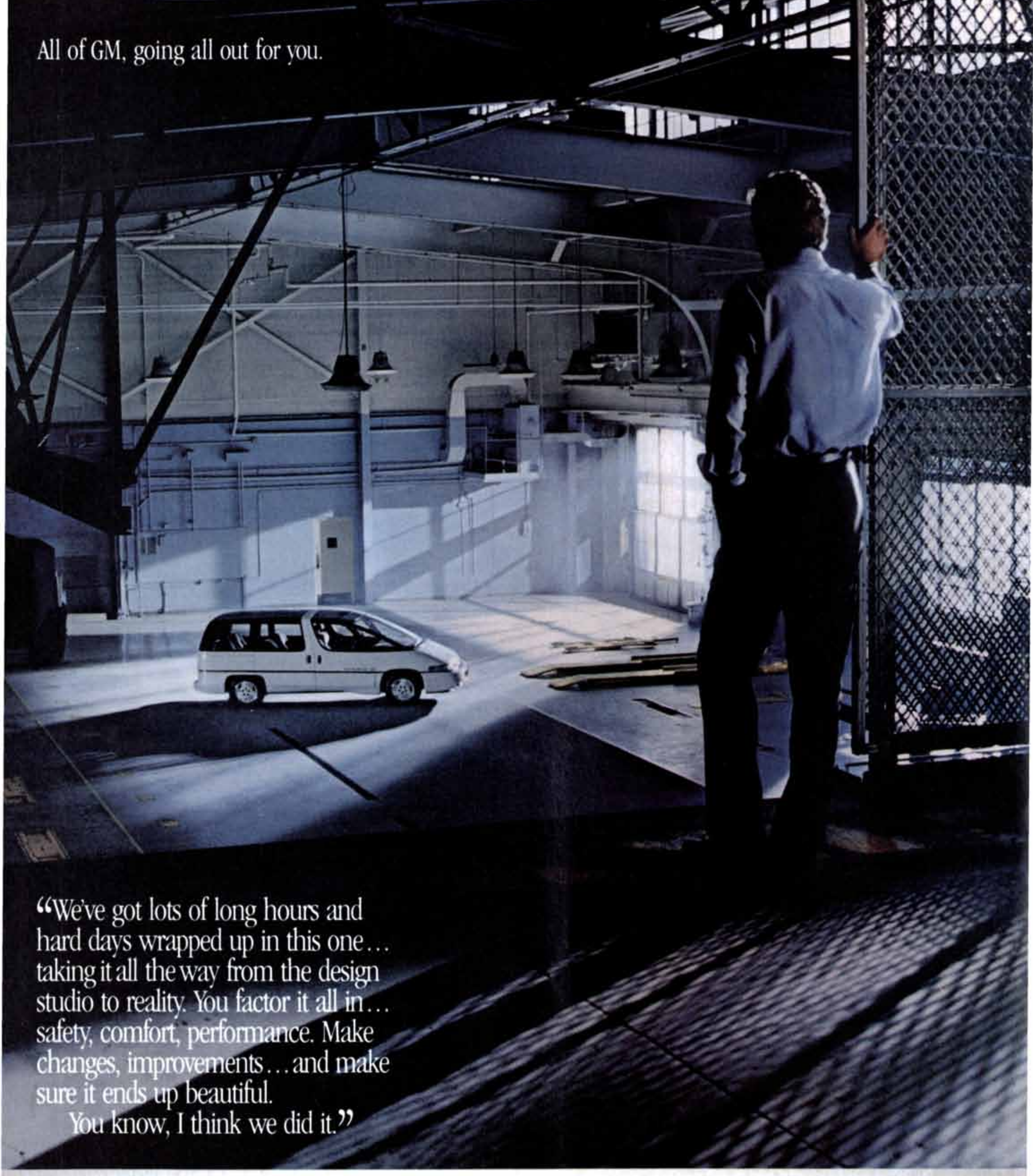
The solution to this problem is you...when you join the fight against illiteracy. So call the Coalition for Literacy at toll-free **1-800-228-8813** and volunteer.

Illiteracy may be an epidemic, but with a little caring from you, we can stop it.

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

Ad Council Coalition for Literacy

© 1989 SCIENTIFIC AMERICAN, INC



Fact: GM has been recognized for design excellence by the Industrial Designers Society of America more than any other U.S. company.

GM
MARK OF EXCELLENCE

1990 Pontiac Trans Sport SE at a GM vehicle development center

Chevrolet, Pontiac, Oldsmobile, Buick, Cadillac, GMC Truck

© 1989 SCIENTIFIC AMERICAN, INC

SCIENTIFIC AMERICAN July 1989 Volume 261 Number 1

Testing Weapons in Space

Would a more permissive weapon-testing regime than is set forth in the Antiballistic-Missile Treaty be in the U.S. interest? Reviewing the rationale of the treaty's provisions helps to answer the question

by Ashton B. Carter

When the U.S. and the U.S.S.R. were working out an agreement to restrict antiballistic-missile (ABM) systems in the late 1960's and early 1970's, American negotiators recognized that a treaty merely banning the deployment of a nationwide ABM system would not be enough. In the absence of additional restrictions—on developing, testing and manufacturing ABM-system components—the Soviet Union could easily produce and stockpile the components for later deployment. Indeed, if the Soviet leadership ever decided to break out of the treaty, the U.S.S.R. could have rapidly established an extensive ABM system. Hence, American negotiators wanted an agreement that prohibited the Soviet Union not only from deploying a nationwide ABM system outright but also from positioning itself to deploy one quickly.

At the same time the American negotiators did not want to restrict ABM research unduly. Many federal government officials wanted the U.S. to continue seeking new and possibly better approaches for defending the nation against nuclear attack. ABM-related research was also needed in order to help the U.S. assess Soviet technology and prepare military responses in the event the Soviet Union ended up deploying a nationwide ABM system. In addition, some ABM technologies were useful for other military purposes. The negotiators therefore had to be careful about where they drew the line separating permitted and prohibited ABM-related activities.

The question of where exactly to draw such a line is once again at the center of Soviet-American arms-control talks; this time it has to do with the development and testing of military technologies that can find potential application in space-based ABM systems. The Reagan Administration, having established the Strategic Defense Initiative (SDI) specifically to intensify the exploration of such technologies, naturally wanted to shift the line toward increased permissiveness. The matter was left unsettled, however. To understand the fundamental conflict that underlies the question, it helps to go back and review the purpose of the original negotiations on limiting ABM systems.

It was clear 20 years ago (as it is today) that no treaty could realistically ban ABM research. For one thing, it is not possible to distinguish between ABM and non-ABM research. For another, it is not possible to verify what goes on inside Soviet laboratories. And besides, the U.S. itself stood to benefit from an active ABM research program of its own. For these reasons American negotiators pressed for limits on the engineering development and testing of ABM components—but not on scientific research—in addition to strict limits on the deployment of ABM systems.

Actually, the development and testing limits had to cover more than just ABM-system components; they had to apply as well to other military systems having a significant degree of technological kinship to ABM systems, such as antisatellite (ASAT) systems and weapon systems for defending against enemy aircraft and short-range battlefield missiles. Modified versions of such systems can have potential application in an ABM defense. Similar concerns held for radars and other key sensors in systems that provide warning of missile attacks, track objects in space, monitor weapon test ranges and collect intelligence. In the absence of limits that apply to the development and testing of such technologies, the Soviets could surreptitiously acquire an ABM capability in the guise of pursuing some other military mission and thereby "leak out," rather than break out, of the treaty regime.

The final product of the negotiations—the 1972 ABM Treaty—explicitly recognizes both the breakout and leakout problems. Article I of the treaty obliges each side not only to refrain from deploying a nationwide ABM defense but also "not to provide a base for such a defense." Article VI forbids giving non-ABM systems "capabilities to counter strategic ballistic missiles or their elements in flight trajectory" and forbids testing them "in an ABM mode." Such general legal prohibi-

ASHTON B. CARTER is professor of public policy and acting director of the Center for Science and International Affairs at Harvard University. He earned a Ph.D. in theoretical physics in 1979 from Oxford University. Since then he has worked at Rockefeller University, the congressional Office of Technology Assessment, the Office of the Secretary of Defense and the Massachusetts Institute of Technology. He is currently a consultant to various U.S. government agencies.

© 1989 SCIENTIFIC AMERICAN, INC

tions, however, can be implemented only after they have been translated into specific technical rules.

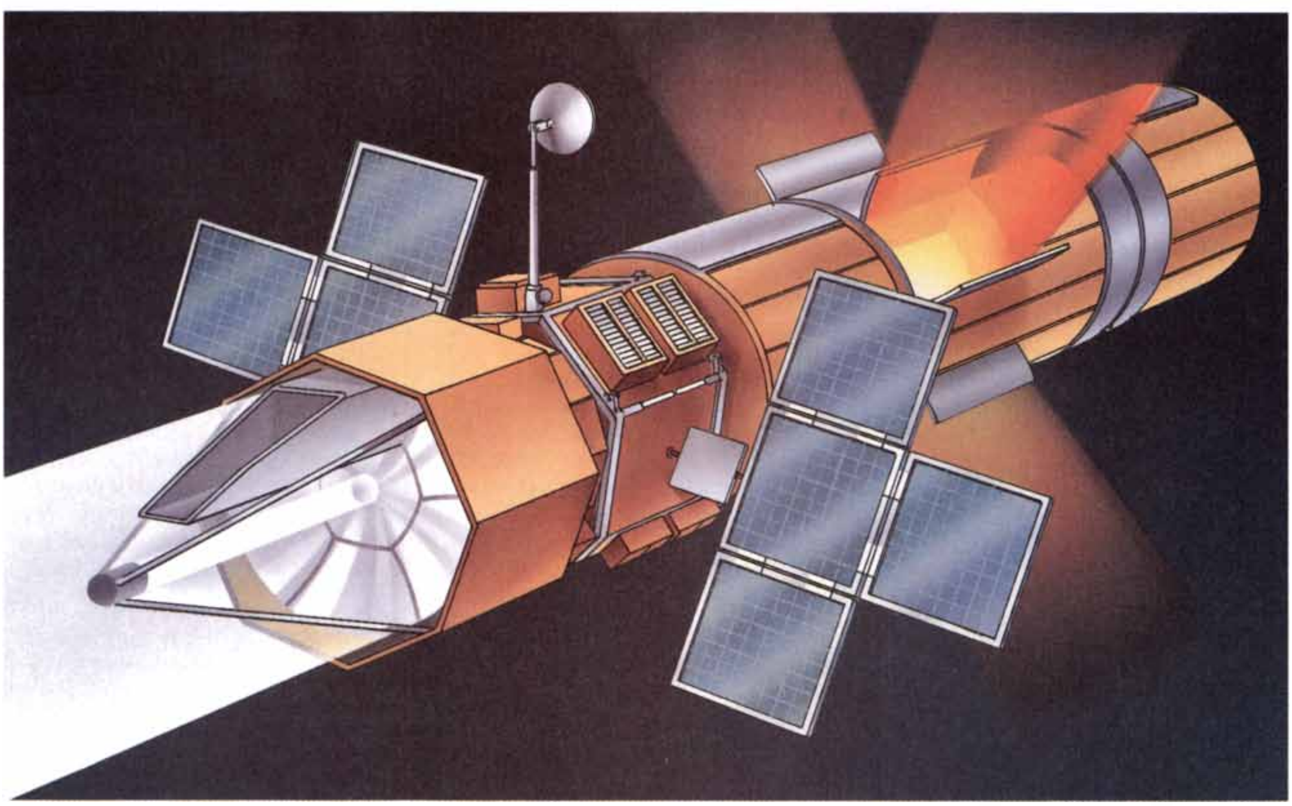
The 1972 treaty (which is still in force) does supply some technical specifics, but most of them apply to the ABM technologies of that era, which consisted primarily of ground-based radars for tracking reentry vehicles and ground-based nuclear-tipped interceptor missiles for destroying them. The negotiators of the treaty were well aware that ground-based radars and interceptors were not the only technologies of potential relevance to an ABM defense. Space-based lasers, particle-beam generators, infrared sensors and most of the advanced technologies currently under intensive study in the SDI program had already been proposed for ABM missions by 1972—and in most cases a good deal earlier. Yet the negotiators believed that detailed questions of interpretation pertaining to such technologies could best be answered if and when they advanced to the point where either side grew concerned about the other's breakout or leakout from the treaty.

The U.S. has not only sought to constrain Soviet breakout and leakout potential in the formulation of technical rules; it has also sought to give its own ABM program enough leeway to collect the data needed to determine whether the U.S. ought to continue abiding by the treaty or deploy an ABM defense. Because one set of rules—applying equally to the U.S. and the U.S.S.R.—must balance the threat of a rapid deployment of an ABM system against the need to explore ABM-related concepts, the two objectives can occasionally come into conflict.

Indeed, such a conflict has arisen in connection with the development and testing of space-based weapon systems. Officials in the Reagan Administration, fearing that the formulation of a new set of interpretative rules could effectively kill the SDI program, refused to enter into technical discussions with the Soviets. Instead they aggressively advocated a "broad interpretation" of the treaty that allowed unrestricted testing of advanced ABM components in space. Soviet officials, for their part, argued that the research of space-based ABM systems should be limited along with their development and testing—a position that is also inconsistent with the original ABM Treaty. President Bush now has to grapple with the issue his predecessor left unresolved.

Resolution of this issue is essential to the continued viability of the ABM Treaty. Designers of ASAT systems and space-based sensors, both in the U.S. and the U.S.S.R., need to know how the treaty's limits apply to their work. Analysts monitoring compliance with the treaty need to know how to recognize a violation. And Congress and the U.S. public need to be reassured that the treaty is being heeded responsibly.

Clarification of the ABM Treaty's limits on advanced military technologies could be pursued during the coming years in formal arms-control negotiations between the U.S. and the U.S.S.R. Yet formal agreements are not necessary: the SDI Organization and its executive- and legislative-branch overseers could instead state informally which types of space experiments they view as permitted and which as prohibited—in effect setting an example for the Soviets to emulate. As long as both sides are satisfied that the other is following a reasonable interpretation of the ABM Treaty, the issue



ZENITH STAR LASER TEST, planned for the mid-1990's as part of the U.S. Strategic Defense Initiative (SDI) program, is meant to evaluate the weapon potential of space-based chemical lasers. In the test an orbiting hydrogen fluoride laser will direct its beam at orbiting targets. Some observers maintain that the Zenith Star test contravenes the 1972 Antiballistic-Missile (ABM) Treaty, which explicitly forbids tests of ABM-system components based in space. The SDI Organization, however, plans to test the laser as an antisatellite (ASAT) weapon, which does not fall directly under the purview of the treaty.

© 1989 SCIENTIFIC AMERICAN, INC

need not stand in the way of other arms-control agreements.

However the clarification of the limits on testing space-based weapons is pursued, it will need to be based on the technical facts and on a sense of the range of alternatives. In spite of its attempt to undermine the traditional interpretation of the ABM Treaty, the Reagan Administration pledged that the U.S. would continue to abide by it. As traditionally interpreted, the ABM Treaty forbids any test of a space-based weapon that entails intercepting a strategic ballistic missile in flight. Such a "full-up" ABM test is illegal whether the weapon is a homing interceptor missile, a high-energy laser, a particle-beam generator or any other type of weapon. The treaty's ban on full-up ABM testing of space-based weapons applies no matter what the technical characteristics of the weapon are: there is no threshold of weapon "strength" below which such tests would be permitted. ABM weapons can be legally tested on the ground, but such testing has to be carried out at a designated ABM test range. In many cases ground-based testing is fully adequate to obtain the technical data necessary to determine a space-based weapon's feasibility.

Nevertheless, the traditional interpretation has loopholes that allow tests of space weapons to be conducted in such a way that many of the technical features of an illegal full-up ABM test can be replicated. In a "lofted mode" test, for example, the weapon that intercepts the ballistic-missile target is not placed in a stable orbit around the earth. Instead it is boosted into a suborbital trajectory so that it can spend a few minutes in space while the test is conducted and then reenter the atmosphere. Although the weapon is in space only briefly, its position and velocity in relation to the target can be nearly the same as in a full-up test.

In fact, the U.S. has announced plans to test ABM interceptor missiles in a lofted mode. Officials of the SDI Organization argue that if the weapon is carried into space by a booster launched from a designated ABM test range it should be considered ground-based and not space-based—even though it is being designed as part of a system to be deployed in space. Because tests of ground-based ABM systems are allowed (even under the traditional interpretation of the treaty), the U.S. can maintain that its planned interceptor test will not contravene the testing limits of the treaty. Yet if this loophole in the treaty remains open, a space-based interceptor can be legally tested in the lofted mode until it has achieved full ABM capability.

Interceptor missiles can readily be tested in the lofted mode, but powerful chemically pumped lasers (which have been proposed as space-based weapons) are too unwieldy and expensive to be boosted into short-lived suborbital trajectories. Still, another loophole in the traditional interpretation of the ABM Treaty can be exploited to test such a laser in space without violating the letter of the treaty. Although the traditional interpretation bans the testing of ABM systems in space, it does not ban the testing of ASAT systems in space.

It is therefore perfectly legal to perform weapon tests in an "ASAT mode," in which an orbiting weapon intercepts an orbiting target. The Soviet Union has tested interceptor spacecraft in that mode for well over a decade, and the U.S. intends to conduct an ASAT-mode test of a large hydrogen fluoride laser weapon in the mid-1990's under the code name Zenith Star. An ASAT-mode test can easily be made to simulate an ABM-mode test by having the target "satellite" fire a rocket thruster during the test so that it mimics a strategic ballistic missile in the boost or post-boost phase.

Weapon tests in the ASAT mode, unlike those in the lofted mode, are limited to a certain extent by the ABM Treaty. The provision against leakout contained in Article VI of the treaty sets an implicit threshold above which ASAT-mode testing would be illegal, because it prohibits endowing non-ABM weapons such as ASAT systems with an ABM capability. The Zenith Star space laser may or may not exceed that threshold: it all depends on what constitutes "an ABM capability."

It is not difficult to imagine how thresholds might in principle be defined for lasers, rocket interceptors and particle-beam generators that are tested in space for nominal ASAT or air-defense systems but that are potentially applicable in ABM systems as well. For example, the rate at which a space-based laser weapon can destroy targets at a given range—and hence its inherent ABM capability—is determined in large part by its brightness: the power of the light emitted by the laser divided by the angular size of the diverging cone into which the laser's pointing mirror can focus the beam. An upper limit on the brightness of lasers in space might therefore serve as a viable constraint on the ABM capability of such weapons.

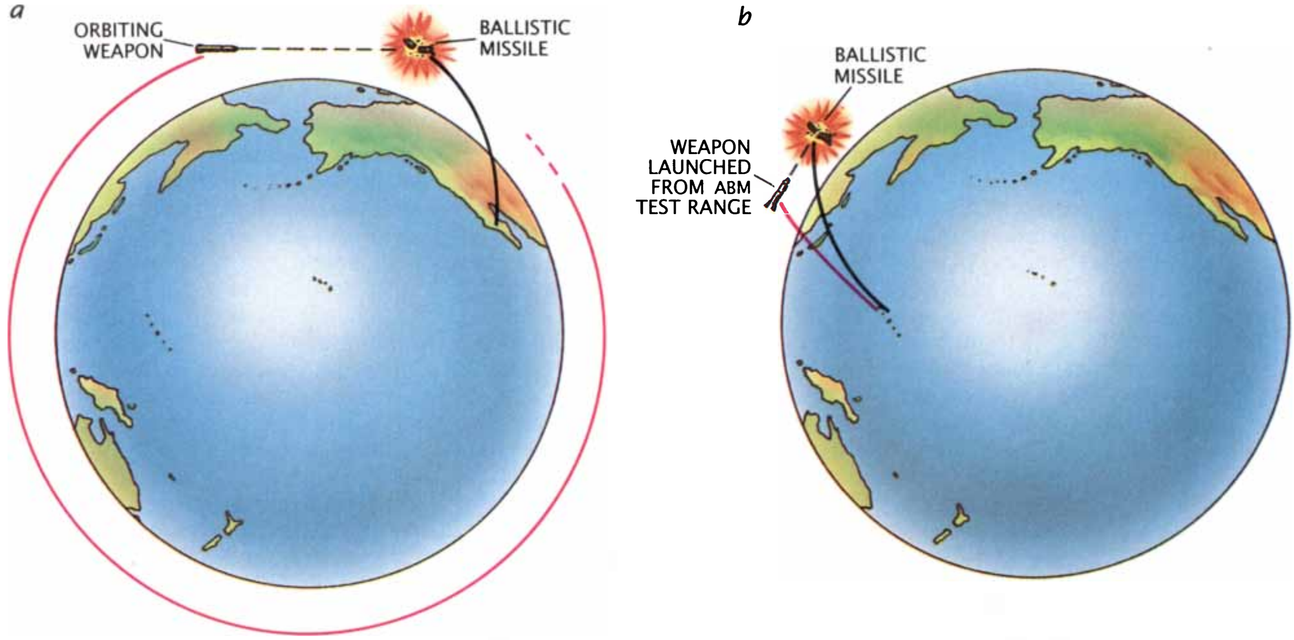
The American Physical Society's Study Group on the Science and Technology of Directed Energy Weapons suggested that even a missile booster "hardened" to withstand laser illumination would suffer relatively rapid structural damage if it were exposed to an incident energy flux in excess of 300 kilowatts per square centimeter. At a typical ABM interception range of 2,000 kilometers, delivery of such a flux requires a laser having a brightness greater than about 10^{22} watts per steradian. Brightnesses of such magnitude are unlikely to be achieved soon; to date the most powerful chemical lasers (which are ground-based) are reported to have a brightness of some 10^{17} watts per steradian.

Certainly if the U.S.S.R. were someday to test lasers of 10^{22} watts per steradian brightness in space, the U.S. would have to worry about the Soviets attaining an ABM capability through breakout or leakout—even if they maintained that the lasers were designed for ASAT systems. Actually, American defense analysts would be concerned if the Soviet Union put lasers having a good deal less brightness in space: the current U.S. strategic nuclear forces, consisting of Minuteman and MX/Peacekeeper intercontinental ballistic missiles (ICBM's) as well as Poseidon and Trident submarine-launched ballistic missiles (SLBM's), are not hardened against laser damage. In addition, a Soviet space-based laser might come within less than 2,000 kilometers of a U.S. missile in some scenarios of potential military concern. Hence, the brightness threshold for concern is closer to 10^{17} than to 10^{22} watts per steradian. Presumably Soviet calculations would lead them to the same conclusion about U.S. lasers in space. Thresholds of ABM capability for other types of weapons can be similarly defined.

Although single-parameter thresholds may be technically elegant and intuitively appealing, in practice they are problematic. The ABM potential of a space-based laser depends not only on its brightness but also on its wavelength, on the supply of laser-pumping energy on board its space platform and on other factors. Moreover, the acquisition, pointing and tracking systems that direct the laser beam to targets are themselves characterized by an array of physical parameters that directly affect a space-based laser's ABM capability.

Another major difficulty in negotiating weapon thresholds is their verifi-

© 1989 SCIENTIFIC AMERICAN, INC



FOUR MODES of testing a space weapon are treated differently by the ABM Treaty, in spite of the fact that they can—to a certain extent—be made to replicate the same test conditions. In a "full-up mode" ABM test (*a*) a weapon in stable orbit intercepts a strategic ballistic missile in flight. In a "lofted mode" test (*b*) the intercepting weapon is launched on a suborbital flight; it is technically not "based" in space. In an ASAT-mode test (*c*) both weapon and target are placed in orbit. In an air-

cation. Each party to the ABM Treaty should be able to assure itself that the other party is complying with the agreement. If brightness thresholds are to be the basis for judging treaty compliance for space-based lasers, each side must have some method of measuring the brightness of the other's test lasers in space.

The U.S. and the U.S.S.R. could deploy a constellation of satellites capable of detecting laser beams, but it would be unlikely for one side's satellites to be in line with the beam of the other side's space laser at the time it was tested. Intelligence sensors would instead have to collect laser energy scattered by the focusing mirror, space junk, the test target or the earth's atmosphere. Further information might come from other types of sensors capable of detecting the combustion products of the chemically pumped lasers (which are released into space) or the electromagnetic emanations of electrically pumped lasers. The physical size of a space-based laser's optics (which determines the solid angle of its beam) could be measured from the ground or from space by making images of the laser assembly while it is illuminated by sunlight, another laser's light or radar.

Although all of these verification methods are technically feasible, none is foolproof and all are expensive. It has taken the U.S. several decades and many billions of dollars to acquire all the sensors it currently employs to monitor Soviet ABM systems and offensive-missile forces. Comparable amounts of time and money might be necessary for verifying the Soviet Union's adherence to new threshold-based rules for the ABM Treaty.

Yet another weakness of thresholds in setting treaty-compliance rules is that a space weapon might be tested below the threshold and yet be modified to achieve a greater capability. If such upgrading turns out to be relatively easy, then threshold-based limitations provide an insufficient "buffer" period in which to respond to a rapid breakout or leakout: the only buffer left is the time-consuming process of deployment itself.

As traditionally interpreted by the U.S., then, the ABM Treaty prohibits full-up ABM tests of space weapons entirely, does not restrict lofted-mode testing and permits ASAT and air-defense testing up to a currently undefined threshold. Clearly, one key task for the U.S. and Soviet technical specialists discussing the treaty's testing rules should be to define some threshold for ASAT and air-defense testing.

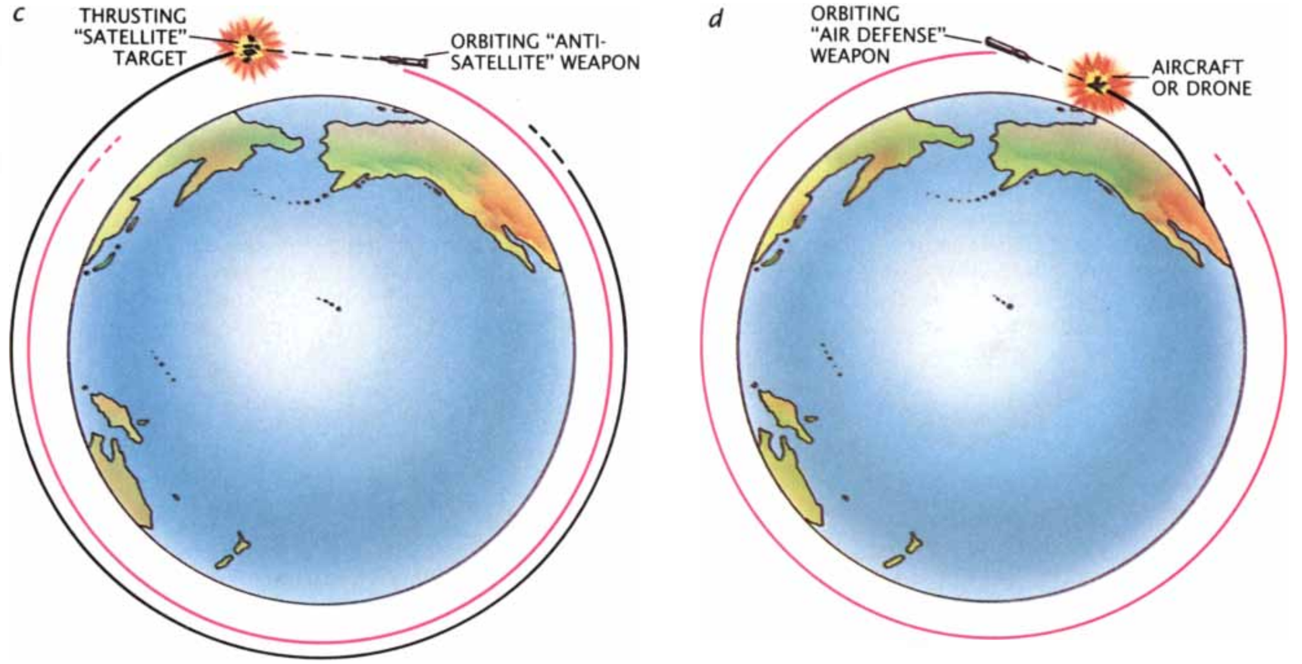
A more important task would be to determine whether ASAT-mode test interceptions of targets equipped with rocket thrusters should be permitted at all. Ordinary satellites, which would be the intended targets of a true ASAT weapon system, fire their thrusters only intermittently and at low thrust levels. Banning ASAT-mode test interceptions of thrusting targets would therefore constrain the development only of ABM systems, not ASAT systems. (Of course the two superpowers might want to limit the development of ASAT weapons, but that would require a new treaty.)

Also, consideration should be given to closing the lofted-mode loophole entirely. The testing of ABM systems at designated test ranges, as allowed by the ABM Treaty, was intended to permit testing of ground-based components only. Yet ABM-capable weapons that are tested in the lofted mode are generally intended for deployment in a space-based, not a ground-based, configuration.

Finally, any agreement clarifying the ABM Treaty's strictures should also include measures that make it easier to verify that both sides are in fact adhering to the agreed limits. In general it is preferable to frame the limits embodied in an arms-control treaty in terms of testing practices rather than numerical thresholds of a weapon's capability, because testing practices are more easily verified.

Rather than simply clarifying and updating the traditional interpretation of the ABM Treaty, U.S. and Soviet negotiators could decide to replace it with new agreements. Such

© 1989 SCIENTIFIC AMERICAN, INC



defense-mode test (*d*) the weapon is aimed at an aircraft in flight or an instrumented target on the ground. The traditional reading of the ABM Treaty forbids the testing of any type of weapon in a full-up mode but does not restrict lofted-mode tests if the weapon is launched from an agreed ABM test range. Weapon testing in space in the ASAT and air-defense modes are also allowed, as long as they do not exceed an unspecified threshold beyond which they are regarded as "ABM capable."

agreements could be either more restrictive or more permissive than the 1972 treaty. In either case, the negotiators would still have to balance the two fundamental goals of the original treaty: to preserve a buffer against the rapid breakout or leakout of an extensive ABM system while allowing the exploration of new technology for ABM defense and other related military missions.

A more restrictive approach might begin with the banning of ASAT-mode weapon tests in space entirely—even against nonthrusting targets. Such an agreement would curtail the development and deployment of space-based ASAT systems as well as space-based ABM systems. (Ground-based ASAT systems and space mines would presumably be still allowed.) It would also obviate the need to define thresholds and thereby ease verification.

Even if the superpowers are unwilling to ban all tests of space weapons, they might at least agree to new and intrusive means of verifying compliance with existing restrictions as contained in the ABM Treaty. In the case of space weapons, an agreement to facilitate verification might require the announcement and description of space tests before they actually take place and even the prelaunch inspection of one another's test payloads.

Another alternative would be to ban nuclear reactors in space. Such a ban would greatly impede the development of many types of space-based beam weapons, because there are no ready alternative sources of energy for the large amounts of power such weapons require. Civilian planetary space probes and low-power nonreactor nuclear-energy sources would be exempt, although they might be subjected to inspection before launch to ensure that they could not serve as part of a weapon system.

At the other extreme, a much more permissive approach to development and testing of weapons in space might be entirely consistent with the ABM Treaty's original purpose. This would be the case if it could be shown that the time required for final deployment provides an adequate buffer against rapid attainment of a space-based ABM capability. In other words, deploying an effective ABM system in space may ultimately require so much time as to make breakout impossible, even if the system's components have already been developed and tested.

In fact, the ABM Treaty embodies such a view with regard to the development and testing of fixed ground-based ABM systems. Tests of such systems are not heavily restricted (as long as they are carried out at agreed test ranges), apparently because treaty negotiators believed the deployment of an effective ground-based ABM system would be so conspicuous and time-consuming that if one side broke out of the treaty and began to deploy such a system the other side would have ample time to respond.

Can a case be made that comparable freedom to develop and test space-based ABM systems would also preserve an adequate buffer against breakout and thereby remain consistent with the ABM Treaty's main purpose? If so, then an agreement that permits extensive space testing while maintaining an absolute ban on deployment could be justified without recourse to the dubious legalistic reasoning of the Reagan Administration's broad interpretation of the ABM Treaty.

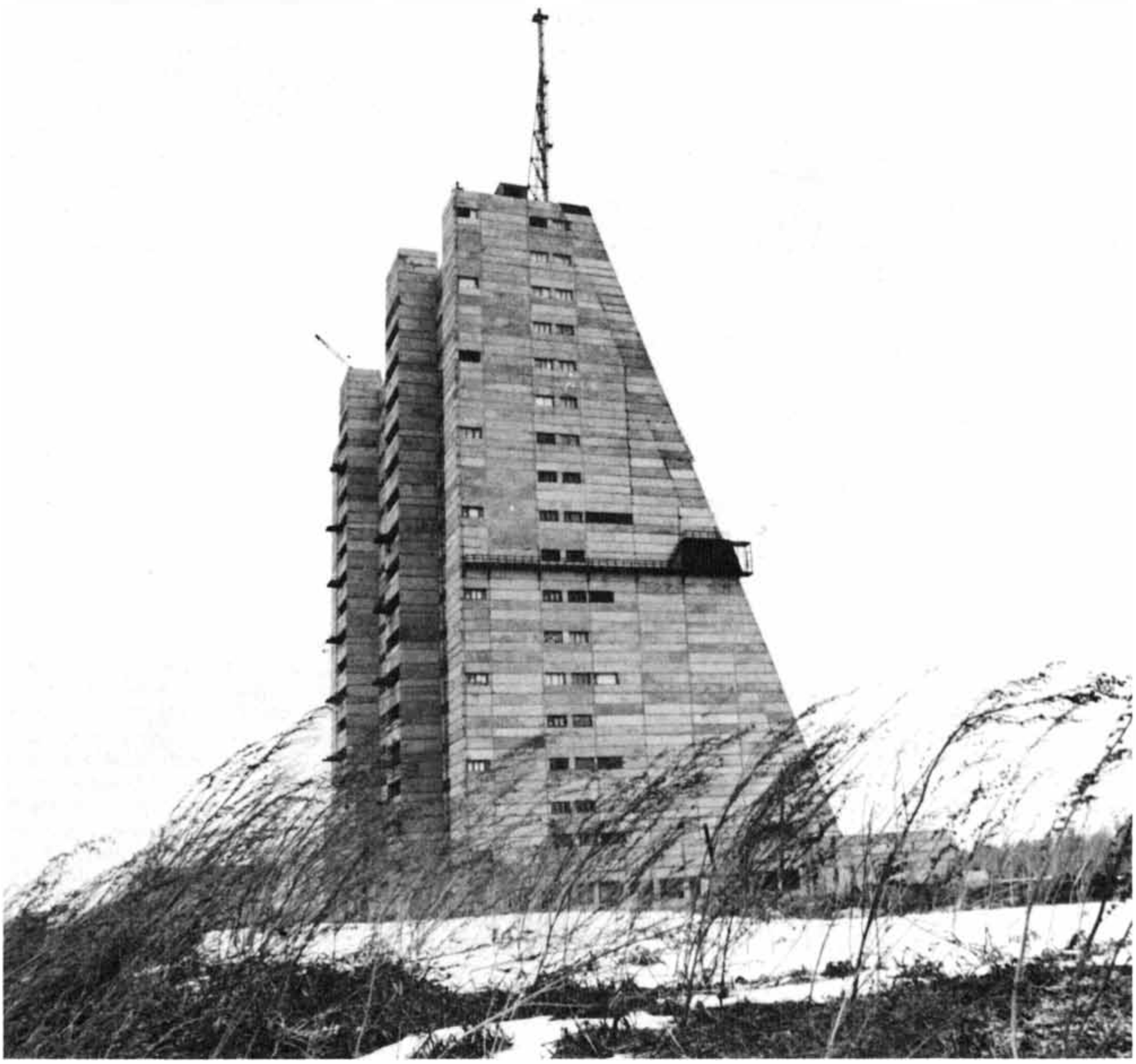
Under such an agreement, each side could be allowed to establish an agreed orbital test range where it could perform tests of space weapons in the full-up ABM mode. There are many ways such a test range could be defined, but the simplest would be for each side to specify the parameters defining a particular orbit: altitude, inclination, longitude and time of ascending node. The orbital test range could consist of a single space structure placed in the specified orbit. All test weapons would then have to be docked with the structure. A more practical scheme would merely require that all weapons involved in ABM tests be clustered together, say, within a spherical volume having a 50-kilometer radius in the specified orbit.

Because the open purpose of such

© 1989 SCIENTIFIC AMERICAN, INC

an orbital test range would be to serve as a setting for realistic testing of ABM-system components, the range should be situated at the altitude where the components would probably be deployed, some 500 to 700 kilometers above the earth. Its orbit would be inclined so that it passed over each side's ICBM and SLBM test ranges and ground-based ABM test ranges. The number of weapons allowed at the orbital test range would be limited, however, so that the range did not itself pose an ABM threat to the other side. All test interceptions in the full-up ABM mode would have to be performed from the range, and no weapon tested at the range could ever be deployed anywhere else. Each side could then station intelligence satellites near the other side's orbital test range to reassure itself that the range posed no breakout threat.

There is little in such an arms-control agreement to hinder the full development and testing of components of a space-based ABM system—by either side. Viewed from the perspective of U.S. negotiators, then, such an agreement would be in the country's national security interest only if it would take a long time for the Soviets to deploy an ABM defense after breaking out of the agreement. During that time the U.S. could beef up its own deterrent forces, deploy ASAT weap-



RADAR AT KRASNOYARSK, rising from its site in the U.S.S.R.'s interior, stands in violation of the ABM Treaty, which requires that all such radars be situated on the country's periphery and face outward. That requirement was intended to make it difficult for either the U.S. or the U.S.S.R. to "break out" or "leak out" of the treaty. In this context, breakout is the rapid deployment of a nationwide ABM system by one side after it has decided to abrogate the treaty; leakout is the surreptitious upgrading of legal ASAT, air-defense or missile-warning systems so that they acquire an illegal ABM capability. Similar breakout and leakout constraints will need to be imposed on future ABM technology in order to maintain the integrity of the treaty.

© 1989 SCIENTIFIC AMERICAN, INC

ons to shoot down the defense system's satellites, build its own ABM defense or enact other countermeasures to nullify the Soviet defense.

Just how quickly could the Soviet Union deploy a space-based ABM system once it had decided to break out of the agreement? The Soviet Union's capacity to launch payloads into orbit establishes a lower limit on the time required for deployment. Unfortunately, this minimum time could be quite short, at least for a first-generation defense of limited capability.

A study by the congressional Office of Technology Assessment supplies a reasonable, if indirectly applicable, standard by which to measure the Soviet Union's breakout potential. The study estimates that the deployment of the first-generation "Phase I" system favored by the U.S. SDI Organization would entail launching payloads with a total mass of between one and two million kilograms into an inclined, low earth orbit. Yet the U.S. Department of Defense predicts that the Soviet Union will be capable of placing one million kilograms per year into low earth orbit by the early 1990's and almost two million kilograms per year by the turn of the century. Almost all of this lift capacity is accounted for by the new Energia launcher, which has an estimated payload of as much as 100,000 kilograms.

If the Soviets were able to launch Energia boosters at the rate foreseen by the Defense Department and if they freely developed and tested components of a space-based ABM defense, then clearly the U.S. would be guaranteed no more than a few years of buffer time against a Soviet breakout deployment of a first-generation space-based ABM system similar to the U.S. SDI's Phase I. For the Soviets the buffer against a U.S. breakout would be much longer, since the U.S. does not have a heavy-lift launcher comparable to the Energia.

Unless U.S. government officials have strong reason to believe that the Soviet Union could not in fact deploy a fully developed ABM defense in a short time, there are only three alternatives that could ensure the U.S. a reasonable buffer period against a Soviet ABM breakout. The first is simply to stick to the traditional interpretation of the ABM Treaty, which effectively hinders the Soviet Union from completing development of space-based ABM weapons by banning tests of space-based ABM-system components.

The second possibility is to negotiate limits on launch capacity along with the establishment of agreed or-

CLARIFY THE TRADITIONAL INTERPRETATION OF THE ABM TREATY

Confirm that the testing of any type of space-based weapon system in a full-up ABM mode is forbidden.

Confirm that simulated or attempted interceptions in space are also banned.

Confirm that the same rules apply to the boost-phase interception of both strategic and nonstrategic missiles.

Determine whether lofted-mode weapon tests are permitted or prohibited.

Determine whether ASAT-mode weapon tests against thrusting targets are permitted or prohibited.

Determine the threshold at which weapons tested in the ASAT and air-defense modes are "ABM capable." Confirm that weapons exceeding the threshold are not allowed to be tested in space.

Determine the threshold at which space-based ASAT and air-defense systems attain an ABM capability or create a base for rapidly establishing an ABM defense. Confirm that the deployment of systems exceeding the threshold is banned.

Agree to forbid specified types of interference with verification (telèmetry encryption, for example) during tests of space weapons in all allowed modes.

Agree to require that tests of space weapons in all allowed modes be carried out from or along customary ICBM and SLBM test ranges and agreed ABM test ranges or from space weapons passing over those ranges.

NEGOTIATE A MORE RESTRICTIVE REGIME

Agree to most or all of the points listed above.

Agree to forbid all lofted-mode weapon tests.

Agree to forbid all ASAT-mode weapon tests against thrusting (and perhaps even nonthrusting) targets.

Agree to forbid all air-defense-mode tests.

Agree to announce and describe tests of space weapons before they actually take place.

Agree to allow prelaunch inspection of specified payloads connected with tests of space weapons.

Agree to ban nuclear reactors in space except those on civilian spacecraft, which would be subject to prelaunch inspection.

NEGOTIATE A MORE PERMISSIVE REGIME

Define orbital test ranges at which each country can carry out its tests of space weapons.

Agree on the number and types of space weapons that can be tested at an orbital test range.

Determine which types of tests are prohibited in space except at an orbital test range.

Agree to certain measures that facilitate monitoring of activities at an orbital test range.

Agree that all test targets are to be launched along customary ICBM and SLBM test ranges or agreed ABM test ranges.

Agree to forbid tests of components of deployed space-based ASAT or air-defense systems in a full-up ABM mode.

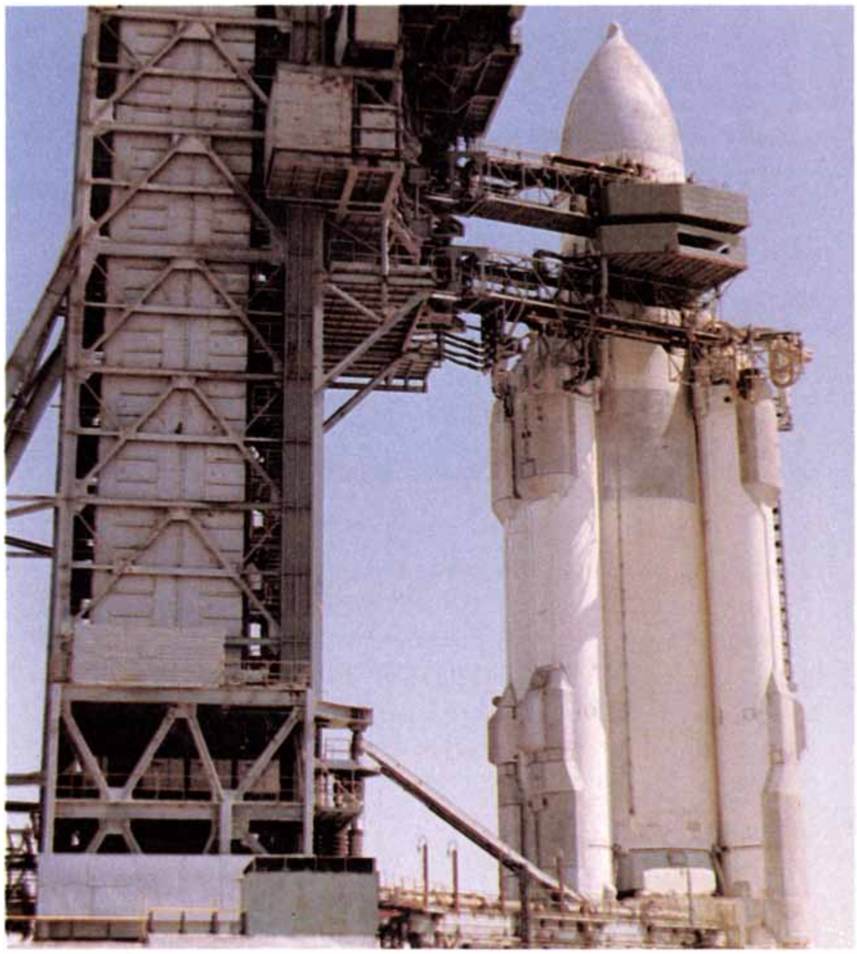
Determine the threshold at which space-based ASAT and air-defense systems attain an ABM capability or create a base for rapidly establishing an ABM defense. Agree to ban deployment of systems exceeding the threshold.

Agree to limit heavy-lift launch capability in order to preclude the rapid deployment of a space-based ABM system.

Prepare unilateral countermeasures as needed to ensure effectiveness of offensive-missile force, given the increased risk of breakout and leakout associated with a permissive weapon-testing regime.

THREE APPROACHES to negotiating an agreement to limit the testing of ABM-related weapons in space differ in the strictness of the limits. The goal is to craft mutually acceptable restrictions that allow a reasonable amount of research and testing of ABM concepts while ensuring to each side a "buffer" period long enough to enact effective countermeasures if the other side breaks out or leaks out of the agreement.

© 1989 SCIENTIFIC AMERICAN, INC



ENERGIA LAUNCHER affects the ABM Treaty's capacity to hinder a Soviet breakout. It is thought that a single Energia launcher can place 100,000 kilograms into low earth orbit. (The U.S.'s most capable launcher—the space shuttle—has a maximum payload of about 25,000 kilograms.) Given that capability, a first-generation space-based ABM system could in theory be deployed by the U.S.S.R. fairly rapidly.

bital test ranges. These limits might focus on the number of launch pads for heavy-lift boosters; the size of launcher-assembly buildings connected to the pads by rail or heavy-duty road; the capability for manufacturing and transporting the enormous rocket stages of a heavy-lift launcher; and the number and size of facilities for producing, transporting and storing liquid-hydrogen fuel.

The third alternative is to compensate for the eroding buffer by deploying (or making preparations to deploy) countermeasures to possible Soviet ABM systems. The U.S., for example, could build more offensive missiles or abandon certain targeting plans that critically rely on the unimpeded passage of nuclear missiles.

In spite of the open loopholes, the undefined thresholds and the dubious interpretations, the U.S. and the U.S.S.R. will probably continue to find it in their national security interest to keep the ABM Treaty regime intact. The reason is simply that neither side currently knows how to build missile defenses that are technically and militarily attractive. Nevertheless, the ABM Treaty's rules on the development and testing of military technology do require periodic revision as the technology changes. It seems likely that such updating, in particular with regard to space weapons, will figure prominently in superpower arms-control negotiations in the coming years.

In discussions of such technical rules and definitions it is important to recognize that there is considerable latitude for drawing the line between permitted and prohibited ABM-related activities. On the one hand, development and deployment of a space-based ABM system relying on current technology would be an inherently conspicuous and time-consuming process. Neither side need worry that the other could break out of the treaty regime and quickly deploy a system that is more than a nuisance. It can be reasonably argued on those grounds that the rules governing the development and testing of space weapons can safely be made fairly permissive.

On the other hand, all of the crucial technical work bearing on the feasibility of advanced ABM concepts can be accomplished for the time being without extensive space testing, within the traditional interpretation of the 1972 ABM Treaty. (Indeed, elaborate space tests of exploratory technology can be a waste of money.) For that reason it can also be reasonably argued that the rules might as well be made fairly restrictive.

There is also no particular urgency to the discussions, since few advanced ABM technologies are truly approaching significant military capability today. Although many technologies will be explored by the U.S. as part of the SDI program, few—if any—of the concepts on today's SDI briefing charts will survive to the stage of serious preprototype development. Comprehensive and precise technical rules covering all types of lasers, particle-beam generators, interceptors and sensors will not be necessary soon.

Nevertheless, a consensus between the U.S. and the U.S.S.R. to pursue seriously the updating of the ABM Treaty would demonstrate that the superpowers understand the technical and military basis of this important arms-control agreement, and it might even help in their negotiations to reduce offensive nuclear weapons. It is probably more important to clarify the line between permitted and prohibited ABM-related activities than to draw it at a certain place.

FURTHER READING

DIRECTED ENERGY MISSILE DEFENSE IN SPACE. Ashton B. Carter. Office of Technology Assessment, OTA-BP-ISC-26, U.S. Government Printing Office, April, 1984.

REPORT TO THE AMERICAN PHYSICAL SOCIETY OF THE STUDY GROUP ON SCIENCE AND TECHNOLOGY OF DIRECTED ENERGY WEAPONS in *Reviews of Modern Physics,* Vol. 59, No. 3, Part II; July, 1987.

THE STRUCTURE OF POSSIBLE U.S.-SOVIET AGREEMENTS REGARDING MISSILE DEFENSE. Ashton B. Carter in *On the Defensive? The Future of SDI.* Edited by Joseph S. Nye, Jr., and James A. Schear. University Press of America, 1989.

DEFENDING DETERRENCE: MANAGING THE ABM TREATY REGIME IN THE 21ST CENTURY. Edited by A. H. Chayes and P. Doty. Pergamon-Brassey, in press.

© 1989 SCIENTIFIC AMERICAN, INC

Corp. 9-104-F ©Monsanto Company

"The whole of science is nothing more than a refinement of everyday thinking."

—Einstein

"I'd give my right arm for a carpet that won't stain."

—Mom

It's a basic science lesson that's largely overlooked: if you want to make an impact with your inventions, don't bury your nose in the laboratory until you've put your finger on the pulse of the consumer.

At Monsanto, anticipating demand and combining it with patentable discovery is not only encouraged, it's rewarded. Monsanto's Edgar M. Queeny Award honors the development of proprietary technology that leads to commercial success and, this year, carries a $160,000 prize plus individual gold medallions.

We're proud to present the 1989 Queeny Award to Randolph C. Blyth, David P. Floyd, Paul E. Sikorski and Pompelio A. Ucci for the discovery and development of Wear-Dated® Carpet's patented, fiber-applied locked-in stain resistance—a quantum leap towards the ultimate goal of a stain-proof carpet. And a breakthrough that Monsanto offers millions of carpet customers. From a net revenue of essentially zero in 1986 to projected sales in the hundreds of millions for

Sikorski *Ucci* *Floyd* *Blyth*

1988, their discovery resulted in the most successful product launch in Fibers Division history.

All due to the ongoing work of four dedicated scientists with enough vision not to overlook our simple, everyday needs.

Monsanto

© 1989 SCIENTIFIC AMERICAN, INC

Memory Storage and Neural Systems

Changes in the electrical and molecular properties of nerve cells accompany the learning that takes place in Pavlovian conditioning. The changes offer lessons for the design of artificial networks

by Daniel L. Alkon

A cartoonist portrays the face of a famous personality with a minimum number of lines artfully arranged in a sketch. The sketch provides just enough visual cues to evoke in the person viewing it a remembered pattern that fills in the gaps. In this sense a human being is a pattern-recognition device created in the course of evolution, and memories are the patterns that are stored.

Pattern storage—which is to say memory formation—is governed by a rather simple rule: pieces of a pattern will become linked together in a memory if the pieces are perceived more or less simultaneously. A pattern is formed and stored when a group of pieces or elements is associated in time. The features of a friend's face, for example, are stored in combination and not individually. A pattern of notes is stored in a melody.

Likewise, it is characteristic of memory patterns that awareness of a small portion can trigger awareness of the whole. A familiar feature of a stranger's face may remind one of the face of a friend; a few notes in a refrain may recall an entire symphonic movement.

DANIEL L. ALKON is chief of the laboratory of molecular and cellular neurobiology at the National Institute of Neurological and Communicative Disorders and Stroke. He was trained as a physician, receiving an M.D. from the Cornell University Medical College in 1969, but his curiosity about human behavior and his interest in the molecular aspects of natural phenomena prompted him to turn his investigations to memory. He has pursued those investigations at the National Institutes of Health since 1970; he spent much of that time at the Marine Biological Laboratory in Woods Hole, Mass. When he is not with his family, Alkon relaxes by playing tennis or visiting the Potomac River.

And it is not only elements within a pattern that are linked together—patterns evoke other patterns. The visual pattern of a friend's face can be linked to the sound pattern of the friend's name and to the scent pattern of the friend's after-shave.

Studies of brain function are beginning to reveal how such links are established. The formation of associative memories appears to involve a sequence of molecular changes at specific locations in systems of neurons. The receptivity of some of the neuronal sites, for example, may be greatly enhanced by the migration of a protein called protein kinase C (PKC). The protein moves from the cytoplasm of a cell to the membrane, altering a neuron's properties so that particular input signals trigger impulses more readily. The firing of neurons thus activated reflects the distribution within each neuron as well as within each neuronal system of those sites that have had their excitability permanently enhanced through memory formation.

Many of the molecular transformations involved in memory formation appear to take place in the neuronal branches called dendritic trees, which receive incoming signals. The trees are amazing for their complexity as well as for their enormous surface area. A single neuron can receive from 100,000 to 200,000 signals from the separate input fibers ending on its dendritic tree. Any given sensory pattern probably stimulates a relatively small percentage of sites on a tree, and so an almost endless number of patterns can be stored without saturating the system's capacity.

The studies my co-workers and I have done at the National Institute of Neurological and Communicative Disorders and Stroke, along with the contributions of other investigators, have also hinted at rules that could be used to design computer-based memory systems: neural networks. In fact, the artificial network we have constructed according to biological rules has been quite successful at recognizing patterns, and the mathematical formulas that govern its operation are in turn providing some clues to the biological workings of memory.

In my laboratory we have investigated memory storage and the molecular nature of associative-memory formation by analyzing, in the marine snail *Hermissenda crassicornis* and in the rabbit, a comparatively simple type of associative learning: Pavlovian conditioning. In Pavlovian conditioning an organism learns to link two discrete stimulus elements, just as Pavlov's dog learned to link the smell of meat with the ringing of a bell. The fact that the dog associated the two phenomena was demonstrated by a learned behavioral response to the bell: salivation.

Pavlovian conditioning is evident in the learning behavior of an enormous range of species. In spite of the diversity of organisms, of behaviors and of the stimuli that can become associated, the quantitative rules obeyed are surprisingly similar. The similarity of the rules implies that the functions of the neural systems underlying associative memory in different species are also quite similar. Indeed, an accumulating body of evidence suggests that mechanisms of associative memory have been conserved over the course of evolution.

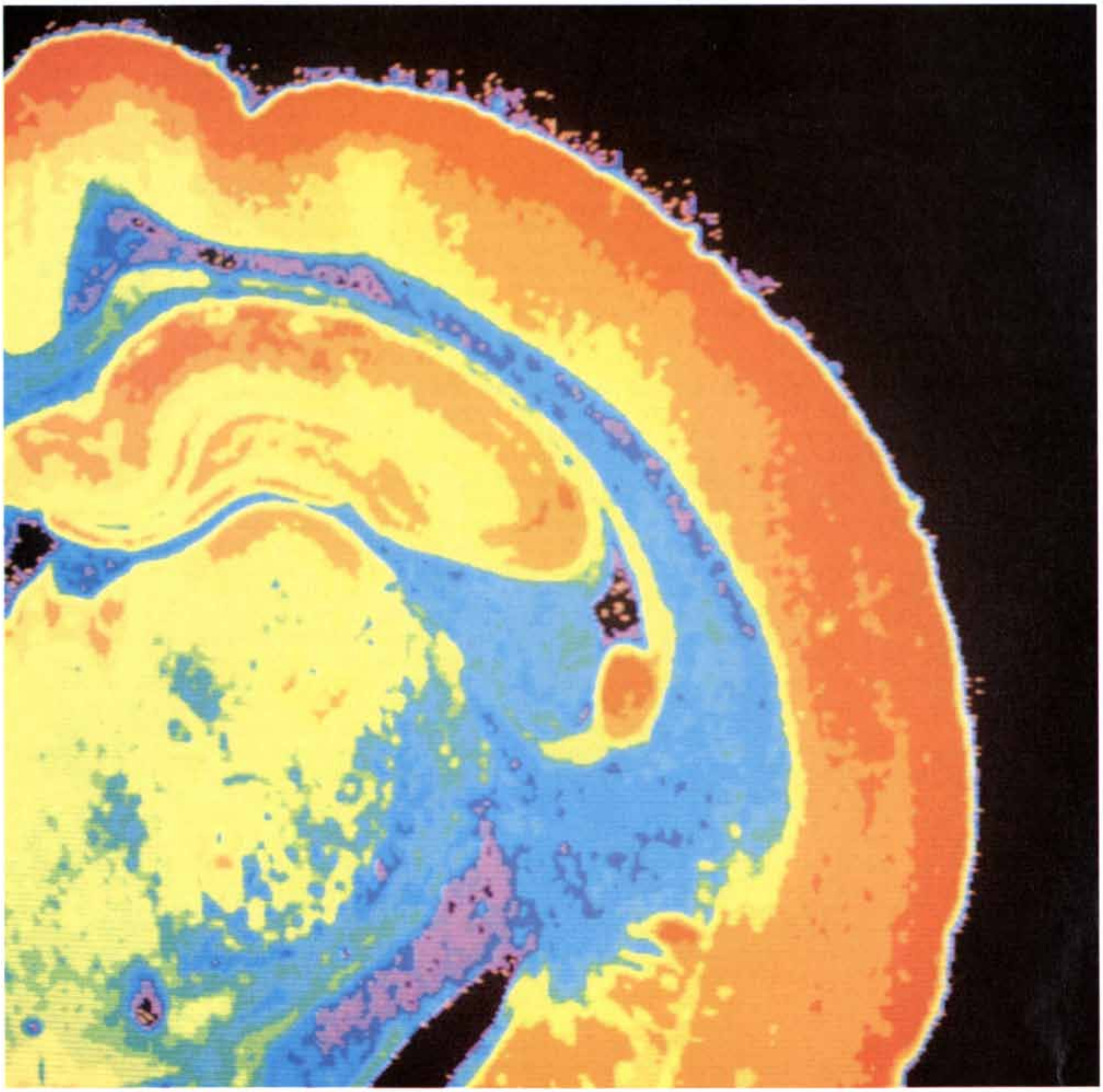
Hermissenda, for example, can be taught to associate the flash of a light with rotation that mimics ocean turbulence; in nature, the snail responds to turbulence by flexing its muscular foot in order to anchor itself to a hard

© 1989 SCIENTIFIC AMERICAN, INC

surface, and through conditioning it learns to do so in response to light [see "Learning in a Marine Snail," by Daniel L. Alkon; SCIENTIFIC AMERICAN, July, 1983]. Rabbits learn to associate an auditory tone with a puff of air to the surface of the eye; the air causes the nictitating membrane to extend, and in time the rabbit extends the membrane when it hears the tone [*see illustration on next page*].

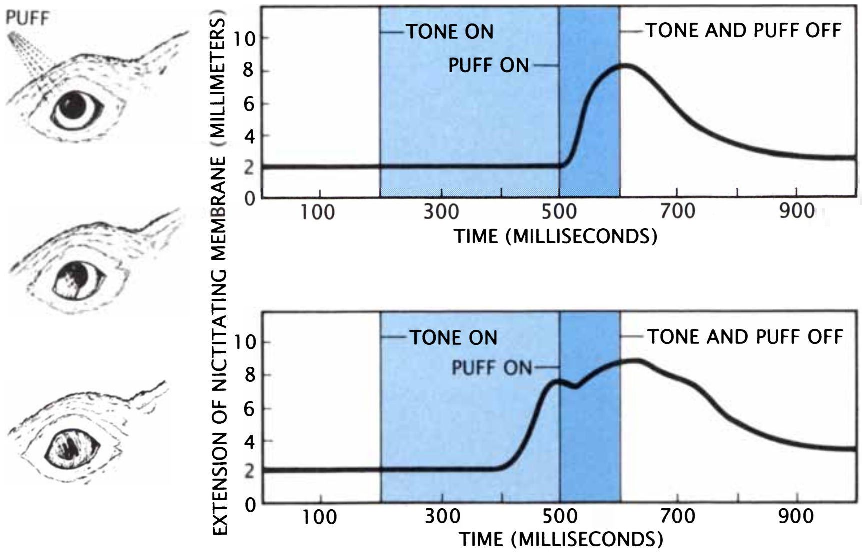
Both of these learned behaviors are examples of response transfer: the transfer of a response ordinarily elicited by an "unconditioned" stimulus (the smell of meat, ocean turbulence or a gust of air) to a "conditioned" stimulus (a ringing bell, a flash of light or a tone). In order to learn the association between conditioned and unconditioned stimuli, an animal must in effect remember the temporal relation of the events. The snail must remember that rotation always accompanies a flash of light, and the rabbit must remember that a puff of air always accompanies the sound of a tone.

Learning and memory in these organisms can be traced to neural systems and to the cellular changes that are responsible for the behavioral changes. In rabbits my colleagues and I have looked at neurons called CA1 pyramidal cells in the hippocampus of the brain; in snails we have looked at neurons known as the Type *B* photore-



CORTICAL SECTION of the rabbit brain was labeled with a radioactive isotope to indicate the amount of an enzyme, protein kinase C (PKC), in or near the membranes of neurons in the section. Red and yellow hues indicate high concentrations of PKC, and blue and violet indicate low concentrations. By comparing labeled sections from trained and untrained animals, the author has been able to identify potential mechanisms for learning and memory. This image was provided by James L. Olds of the author's laboratory at the National Institute of Neurological and Communicative Disorders and Stroke.

© 1989 SCIENTIFIC AMERICAN, INC



TRANSFER OF BEHAVIORAL RESPONSE in the rabbit occurs as a result of associative, or Pavlovian, conditioning. In this case the animal is taught to associate an auditory tone with a puff of air to its eye. The behavioral response—extension of the nictitating membrane (*left panel*)—is transferred from the unconditioned stimulus (the puff of air) to the conditioned stimulus (the tone). The graphs show that before conditioning the membrane extends after the puff of air (*top right*); about 70 trials later, the animal learns to extend the membrane when it hears the tone (*bottom right*). Bernard G. Schreurs of the author's laboratory supplied the data in this figure.

ceptors, which detect light. In both the rabbit and the snail the repeated temporal association of stimuli over the course of Pavlovian conditioning causes a persistent change in these target neurons: the flow of potassium ions through channels in the membranes is reduced.

It is potassium-ion flow and the concomitant flow of other ions that enable nerve cells to conduct electrical impulses. Ordinarily, potassium-ion flow is responsible for keeping the charge on a cell membrane well below the threshold potential at which propagating signals are triggered. When the flow of potassium ions is reduced, impulses can be triggered more readily. Indeed, the excitability of CA1 cells and Type *B* photoreceptors is significantly enhanced by conditioning.

The potassium-ion flow is not reduced in control animals that are not exposed to the stimuli, but neither is it reduced if the same pairs of stimuli are alternated or presented at random. It is not the stimuli themselves, then, that reduce the ion flow, but the temporal relation of the stimuli [*see illustration on page 46*].

The reduction of potassium-ion flow, which is said to change the "weight" of electrical signals within the snail and rabbit neural systems, lasts not just for seconds, minutes or hours but for at least many days and probably much longer. This phenomenon represents a new time domain for membrane-channel function, a time domain never before encountered in fully differentiated (mature) cells. It is a domain uniquely suited for storing associations between stimuli.

In the snail, where neurons are few and their connections are well characterized, it has been possible to establish that the potassium-ion current changes wrought by conditioning are a major cause of the storage and of the potential for recall of the learned association. Such causal implication has not been attempted in the rabbit, although John F. Disterhoft and Douglas A. Coulter then in my laboratory showed that the current changes are specific to learning and are localized within the CA1 cells.

In both the CA1 pyramidal cells of the rabbit and the Type *B* photoreceptor cells of the snail, the ion-flow changes seem to result from the movement of the calcium-sensitive enzyme PKC. In response to the changes in calcium-ion concentration and in another second messenger, diacylglycerol, that accompany the association of temporally related sensory stimuli, PKC moves from the cell cytoplasm to the cell membrane, where it reduces potassium-ion flow.

The translocation and activation of PKC can be induced artificially by the application of a drug called phorbol ester. Other investigators had shown that exposure of CA1 pyramidal cells to phorbol ester causes the enzyme to move to the cell membrane and causes the same potassium-ion flow reduction that takes place in conditioning. Pavlovian conditioning also produces a clear increase in membrane-localized PKC activity and a complementary decrease of activity in the cytoplasm. Barry Bank in our laboratory noted the translocation on the days after the rabbits had been trained [*see top illustration on page 48*].

Similar observations implicate the prolonged translocation of PKC in associative memory in the snail. Exposure of *Hermissenda* Type *B* cells to phorbol ester, together with intracellular elevation of calcium ions, specifically mimics the biophysical consequences of conditioning, that is, it produces the same reduction of potassium-ion flow. The location of PKC in the Type *B* cells strictly determines its effect on potassium-ion flow. In the cytoplasm, PKC increases potassium-ion flow, thereby reducing the excitability of the Type *B* cells; in the membrane, PKC reduces potassium-ion flow, thereby increasing the cells' excitability. Agents that block the translocation of PKC also block the reduction of potassium-ion flow that is ordinarily produced by associative conditioning.

Biochemical evidence that this molecular control is activated by conditioning was provided by measurements in *Hermissenda* of the protein targets for PKC. The enzyme acts by adding phosphate groups to protein substrates. One such target, a protein having a molecular weight of approximately 20 kilodaltons, shows a change in the number of its phosphate groups as a consequence of conditioning. Exposure of the Type *B* cells to phorbol ester also results in the addition of phosphate groups to this protein.

Recently my colleague Thomas J. Nelson and I obtained evidence that the protein is a so-called GTP-binding protein that may be involved in regulating ion channels; when we inject it into Type *B* cells we see a reduction of potassium-ion flow similar to that found after conditioning. This 20-kilodalton protein may serve functions in the context of learning that are analogous to the functions so-called G proteins assume in developmental and oncogenic contexts.

Other observations in both the snail and the rabbit suggest that a second enzyme may contribute to the reduction of potassium-ion flow in Pavlovian conditioning. The enzyme, anoth-

© 1989 SCIENTIFIC AMERICAN, INC

er calcium-activated kinase known as CAM kinase II, also phosphorylates the 20-kilodalton protein target in *Hermissenda* and reduces potassium-ion flow in the presence of elevated intracellular calcium-ion levels. Bank in our laboratory and Robert J. DeLorenzo and his colleagues at the Medical College of Virginia found that the activity of this enzyme increases in the regions of the CA1 cells on days after rabbits have been conditioned but not after the animals have experienced control procedures. The enzyme is concentrated at the so-called postsynaptic (signal-receiving) sites throughout the dendritic trees of the CA1 neurons.

Activation of both the kinases may achieve a more lasting reduction in potassium-ion flow than would activation of either one alone. Cooperation between these enzymes has already been implicated in important physiological effects such as platelet aggregation, insulin secretion and muscle contraction, and PKC translocation is known to account for the prolongation and enhancement of responsiveness to electrical, chemical and hormonal signals in a variety of physiological contexts. Such a ubiquitous role for PKC underscores the notion that a fundamental mechanism for extending the time domain of nerve-cell function has been evolutionarily conserved.

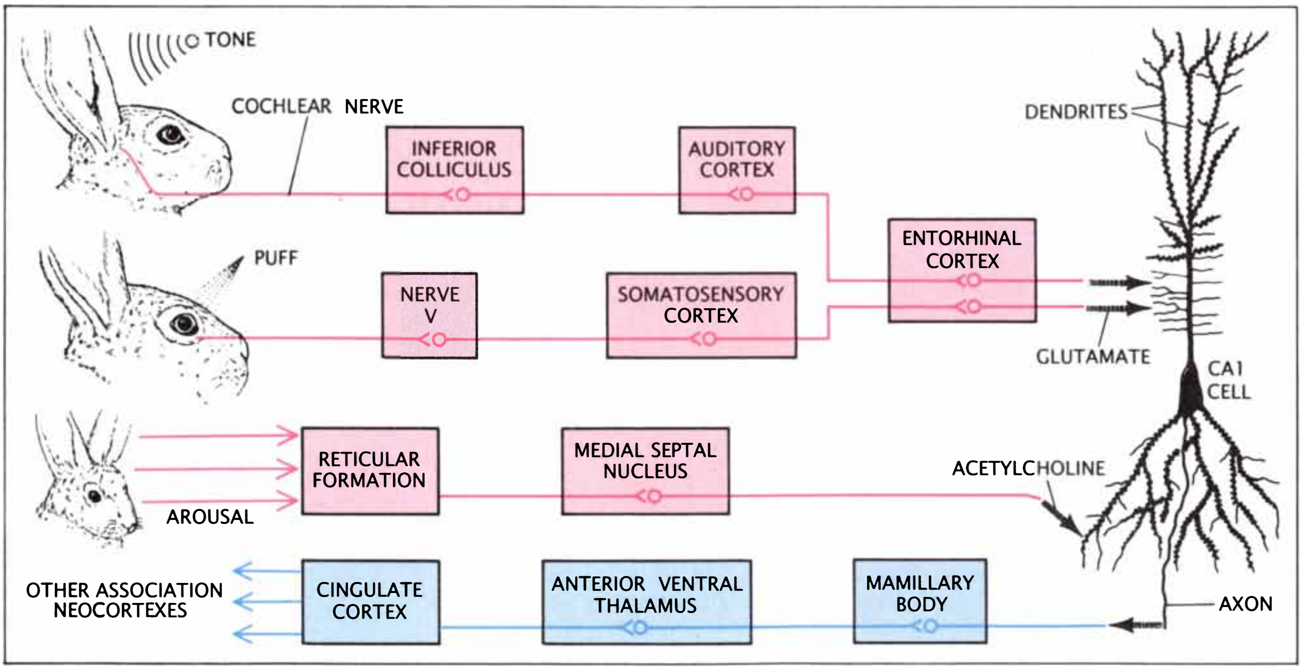
The PKC pathway may be particularly well suited for memory storage because it seems to have the potential to trigger cellular change in even longer and ultimately permanent time domains. It may do so by altering protein synthesis. Phorbol ester–induced translocation of PKC substantially alters the synthesis of a number of neuronal proteins in *Hermissenda,* and the alteration profoundly affects the calcium-stimulated reduction of potassium-ion flow that occurs as a consequence of PKC translocation.

We have also been able to relate the learning behavior of *Hermissenda* to the protein metabolism of the neurons that have been functionally implicated in memory storage. There is a close correlation of memory storage with the cellular quantity of several proteins, one of which is the 20-kilodalton PKC substrate. Nelson in my laboratory found that efficiency of memory storage is also closely correlated in the snail's eye with increases in the synthesis of a number of species of mRNA, a molecular precursor of protein. Again, it appears that one of the mRNA species corresponds to the 20-kilodalton *Hermissenda* protein.

Alterations in the structure of the Type *B* cell branches accompany the changes in protein synthesis wrought by associative-memory storage in *Hermissenda.* The changes can be defined by passing a dye through a microelectrode inserted into the cell body. Five days after snails have been exposed to a training or control experience, the branches of the Type *B* cells in trained snails appear to encompass a much more condensed volume than do the branches of the Type *B* cells in control animals [*see illustration on page 49*]. The magnitude of branching volume is unequivocally related to the magnitude of potassium-ion flow reduction.

The nature of this structural change suggests a hypothesis. It is possible that those branches providing for the synaptic interactions that mediate the newly learned light-rotation association are retained or increase in number, whereas those mediating alternate responses to the light stimulus are eliminated or reduced in number. We are currently testing this hypothesis by injecting cells that have known synaptic interactions with dyes of different colors and then counting the number of synaptic contacts made by Type *B* cells with cells known to mediate different responses.

It is true that animals maintained or reared in stimulating environments tend to have more branches on cortical neurons than do animals exposed to minimum sensory input. But the focusing observed in *Hermissenda* as

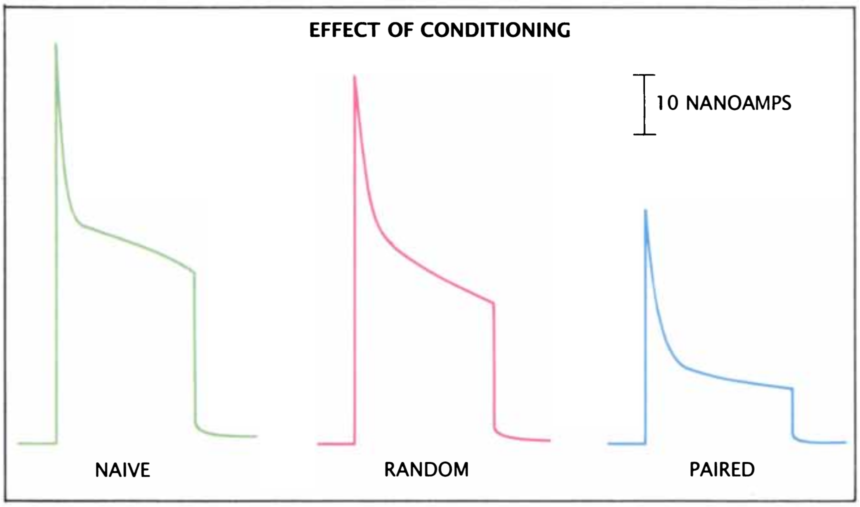


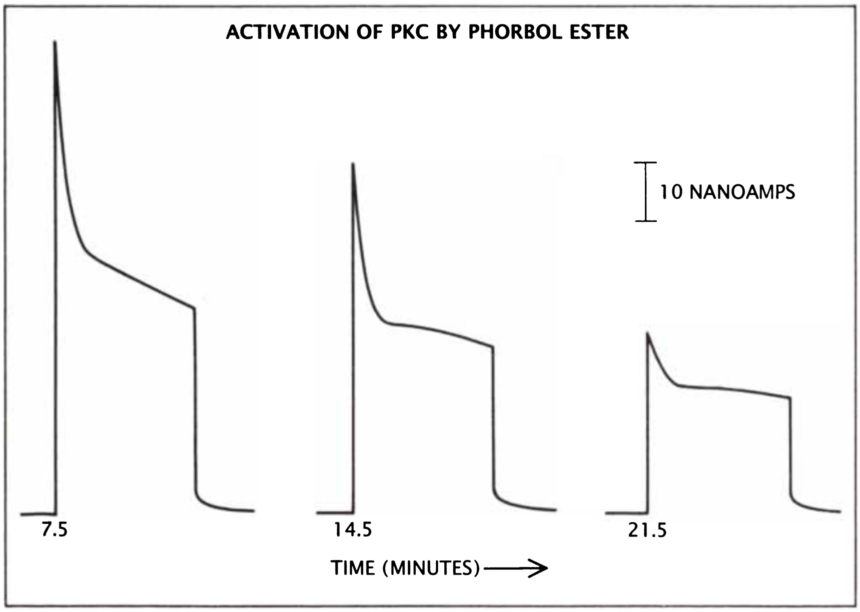
NEURAL PATHWAYS involved in conditioning converge on the CA1 pyramidal cells in the rabbit hippocampus. Impulses from the eye and the ear interact with the CA1 cell by way of the neurotransmitter glutamate; the general arousal of the animal, which has some bearing on how well it learns, is communicated to the CA1 cell by way of the neurotransmitter acetylcholine. The CA1 cell in turn sends impulses to other parts of the cortex. James L. Olds helped the author to design this illustration.

© 1989 SCIENTIFIC AMERICAN, INC

a result of associative-memory storage is quite different from the structural changes measured in training paradigms that are not associative. The Type *B* changes cannot, therefore, be due to the sensory stimulation itself but must be due instead to the timing of the stimuli—the pattern of stimulation.

These findings suggest that a pattern of stimulation is represented and stored by a physical pattern of branching and of synaptic contacts, just as it is represented and stored by a pattern of electrical signals and a pattern of molecular activation. Elimination or reduction of synaptic contacts is also found in the context of development, when multiple neurons compete to establish synaptic contacts with a common target neuron. Jean-Pierre Changeux of the Pasteur Institute in Paris and Gerald M. Edelman of Rockefeller University have extrapolated from the concept of development by postulating that such "neural Darwinism" could take place in the setting of learning and memory. The focusing of Type *B* photoreceptor cells in *Hermissenda* provides a biological justification for the investigators' extrapolation.





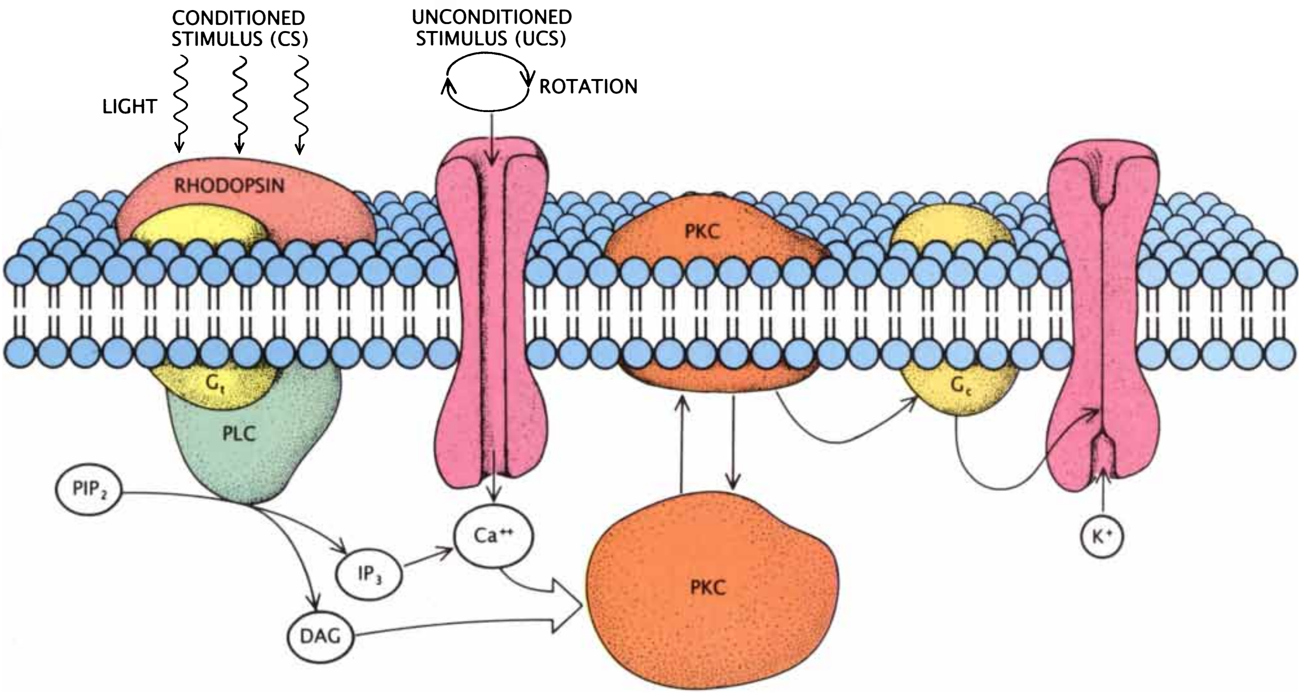
POTASSIUM-ION CURRENTS are reduced in the Type *B* photoreceptor cells of marine snails that have been taught to associate a flash of light with rotation; the animals flex their "feet" in response. There is a marked decrease in the potassium-ion currents in animals that have been conditioned by the paired presentation of light and rotation (*top*); potassium-ion currents are higher when the stimuli are presented at random or when no stimuli are presented at all ("naive"). Activation of PKC in the Type *B* cells with phorbol ester duplicates the effect of conditioning (*bottom*). Phorbol ester causes PKC to migrate to the membrane of nerve cells, and so it is the presence of PKC in or near the membrane that is thought to bring about the reduction of potassium-ion currents. Similar effects are seen in the CA1 cells of rabbits.

In tracing the formation of an associative memory, I have proceeded from stimuli in the environment, to electrical signals in neural systems, to the flow of ions across cell membranes, to molecular controls regulating that flow, to alterations in the synthesis of proteins and finally to changes in the structure of neuronal architecture. It is clear from this progression that the neurons involved in memory storage are dynamic. As mature, differentiated cells they are no longer capable of dividing, yet they are capable of dramatic transformations. In *Hermissenda,* we have shown that these transformations occur in different time domains lasting from seconds to days and longer, and they affect different spatial domains within the cells.

It seems that the neurons in the rabbit hippocampus can also exhibit changes in different cellular spatial domains. We were able to track those changes with a molecular probe first employed by Solomon H. Snyder, Paul F. Worley and their colleagues at Johns Hopkins University. These investigators found that the distribution of PKC within brain structures can be measured by labeling the enzyme with radioactive phorbol ester. If the concentration of the phorbol ester is kept low enough, the label does not cause translocation of PKC and tags only those regions of neurons and neuronal populations that have increased amounts of PKC in the region of the cell membrane.

James L. Olds in our laboratory recently obtained evidence that one day after a training experience membrane-associated PKC increased maximally near the region of the CA1 cell bodies and to a lesser extent in the region of the dendrites (the branches that receive incoming sensory inputs). His findings further suggest that, three days after conditioning, distribution of the PKC label was entirely transformed: labeling is enhanced in the dendritic regions to a much greater degree than it is in the cell bodies. In other words, as the temporal domain of memory storage extended from one to three days, the spatial domain of

© 1989 SCIENTIFIC AMERICAN, INC



MEMBRANE INTERACTIONS during associative conditioning encourage the translocation of PKC. In this schematic diagram of the membrane of a snail Type *B* cell, stimulation by light and rotation initiates a chain of events culminating in the movement of PKC from the cell's cytoplasm to the cell membrane, with a subsequent reduction in the flow of potassium ions through membrane channels. The author suspects that a so-called G protein mediates the reduction of potassium-ion flow.

PKC localized in the region of the membranes may move from the cell body to the dendrites.

The changing spatial domain of PKC distribution may provide some clues to a more general question regarding memory storage. Our own findings as well as those of Richard F. Thompson of the University of Southern California and of Theodore W. Berger of the University of Pittsburgh indicate that the storage of discrete associations in the rabbit seems to involve many cells, a notion that runs counter to our intuition that only a small number of cells would take part. How is it that so many CA1 cells can be changed by a single conditioning event and yet retain the potential for storing many more associations?

It might be that input signals restricted to only small compartments of the CA1 dendritic trees are capable of altering the potassium-ion currents, PKC distribution, protein synthesis and so on in the CA1 cell bodies. The activated cell bodies could then increase the transport of crucial molecules into all the main branches of the dendritic trees, but the molecules would either become localized or exert an effect only in those small compartments that had initially received the appropriately timed input signals [*see bottom illustration on next page*].

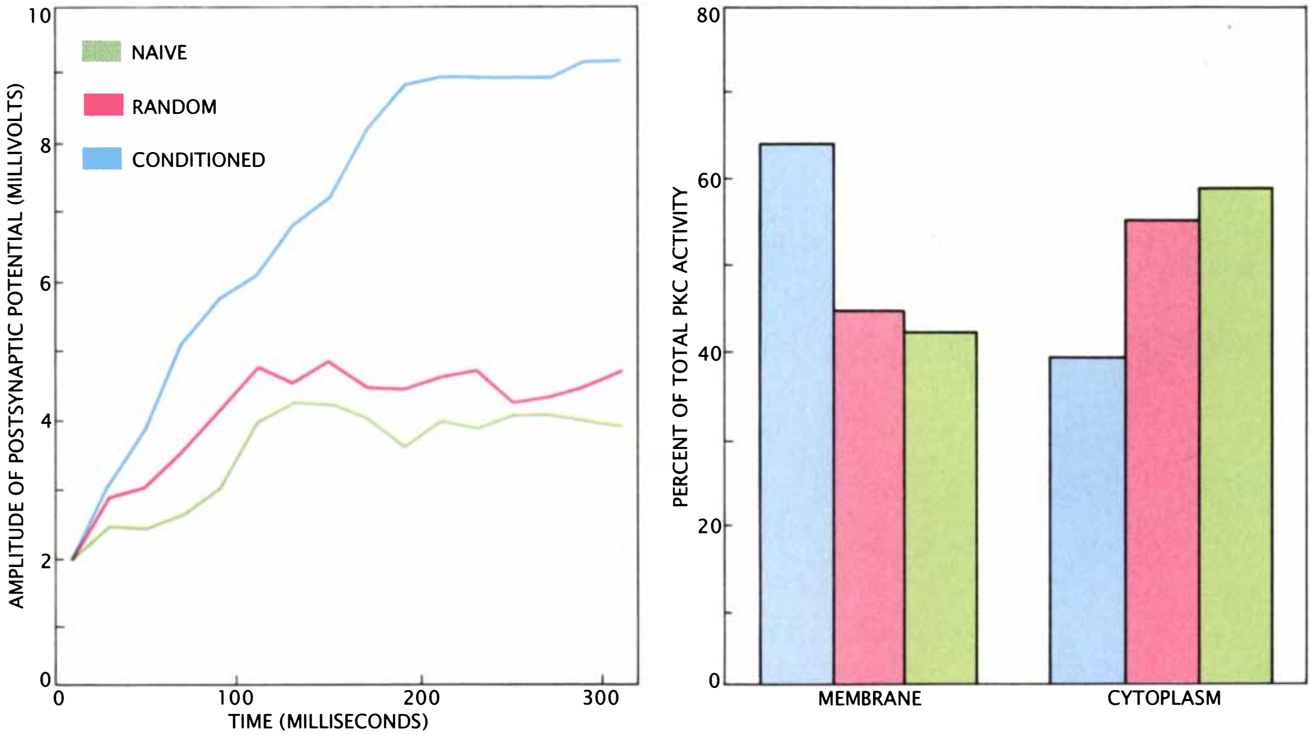
This sequence of events would explain why so many (from 50 to 60 percent) of the CA1 cell bodies in conditioned animals initially show conditioning-induced biophysical and biochemical changes. If an association is eventually stored in particular dendritic compartments (a scenario that is suggested by the distribution of PKC in rabbit neurons three days after conditioning), then specificity of memory storage could be preserved without saturating the capacity of CA1 cells to store additional information.

The working conceptual model I have arrived at through investigations in my laboratory runs somewhat counter to assumptions that have been made about the nature of memory storage. Four decades ago a pioneering student of memory, Donald O. Hebb of McGill University, proposed that memory storage requires that input signals from a single presynaptic source occur together with activity or firing in the postsynaptic element. Because it stipulates activity or firing in the postsynaptic site, Hebb's model suggests that the entire neuron participates in each storage event. Firing of the postsynaptic cell (coincident, in Hebb's model, with presynaptic firing) should affect most or all of the dendritic compartments.

In contrast, what we observe in *Hermissenda* and what we infer from studies of the rabbit hippocampus lead us to believe that there is extensive local interaction between postsynaptic sites. The spread of electrical and possibly chemical signals from one postsynaptic site to another—without activity or firing in the sites—seems to be critical for initiating memory storage. Local storage mechanisms make more sense from a physiological standpoint, since each neuron can store many thousands of memories if the critical interactions are compartmentalized on the dendritic tree. Charles D. Woody of the University of California at Los Angeles has also found evidence for postsynaptic changes in the cat.

A parallel role for local interaction has more recently been implicated in long-term potentiation (LTP), an electrically induced neuronal modification that has served as a model for the alterations induced by natural stimuli during learning. A number of workers, including Holger Wigstrom and Bengt Gustafsson of the University of Göteborg in Sweden, Per O. Andersen of the University of Oslo in Norway, Thomas

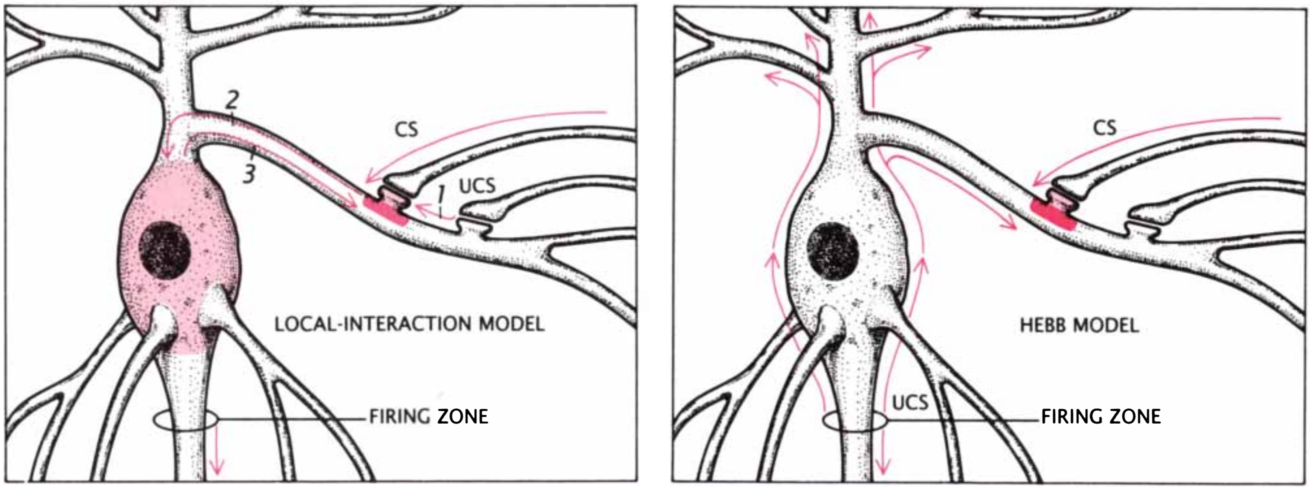
© 1989 SCIENTIFIC AMERICAN, INC



ELECTRICAL AND MOLECULAR PROFILES of the CA1 neurons from conditioned and control rabbits show distinct differences. There is a significant correlation between conditioning and the summated amplitude of the postsynaptic signals generated in the neurons (*left*); summation of the signals with time is greater in conditioned animals. The percentage of PKC in the region of the membrane and cytoplasm of the CA1 cells (measured by activity of the enzyme) also differs in conditioned and control animals (*right*); much more of the enzyme is associated with the membrane in the conditioned animals.

H. Brown of Yale University, Roger Nicoll of the University of California at San Francisco and Gary Lynch of the University of California at Irvine and their colleagues, have studied these local interactions during transient changes in the nerve cells lasting from one to two hours that have become known as associative LTP.

The neuronal loci for changes induced during nonassociative LTP are apparently both presynaptic, as suggested by the work of Timothy V. P. Bliss and his colleagues at the National Institute for Medical Research in London and of Aryeh Routtenberg of Northwestern University, and postsynaptic, as demonstrated by Wigstrom, Gustafsson, Lynch, Andersen and others. Other investigators, notably Eric R. Kandel and his colleagues

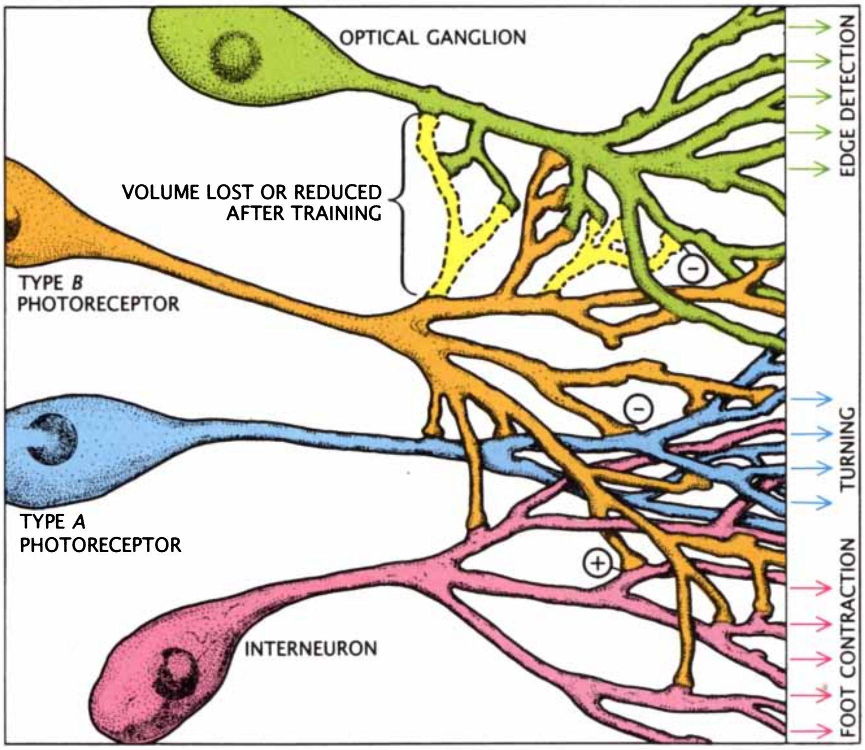


MODELS OF LEARNING-INDUCED CHANGES vary with respect to the interactions necessary to bring about the changes. The local-interaction model, which the author's research supports, postulates that the changes (*color*) originate at a neuron's receiving site when an input there from a conditioned stimulus (CS) is temporally associated with the input from an unconditioned stimulus (UCS) on an adjacent receiving site (*1*). The interaction is communicated to the cell body (*2*), generating factors that return to the site of the interaction and "hard-wire" it (*3*). In contrast, the Hebb model assumes that the changes occur when input from a CS arrives at a receiving site at the same time the neuron is firing in response to a UCS.

© 1989 SCIENTIFIC AMERICAN, INC

at Columbia University College of Physicians and Surgeons, emphasize a presynaptic locus for the nonassociative-memory storage that might occur during, for example, habituation or sensitization.

Many of the questions such research has raised should be solved as molecular PKC probes achieve increased cellular and subcellular resolution. At that time subtle learning-induced differences in PKC distribution should also become evident, whether they take place within intracellular compartments of the neuron, within branches of the dendritic tree or within entire neuronal systems.



"FOCUSING" of the neuronal branches involved in conditioning may eliminate or reduce the volume of the neural pathways that do not link the conditioned stimulus and the behavioral response. This schematic model shows an area of the snail Type *B* photoreceptor cell where reductions of neuronal-branching volume may occur.

Precise mathematical descriptions of the specific images generated during memory storage might help in the assessment of theoretical models of how neural systems within the brain acquire and store information. Such models are also being incorporated in computer-based "neural networks." Already the phenomena we and other workers have observed are constraining and motivating the construction of these artificial models of memory storage by neural systems.

My colleagues Thomas P. Vogl and Kim L. Blackwell of the Environmental Research Institute of Michigan and I thought the conserved-memory mechanisms in the snail and the rabbit might be usefully imbedded in such an artificial system. For example, our observations had indicated that memory storage begins when training stimuli elicit electrical signals that are precisely timed to converge and interact locally at critical neuronal sites such as the Type *B* photoreceptor cell.

This observation suggested the first principle of design for a theoretical associative-learning network: the weight of synaptic interaction, that is, the transfer of signals between elements of the artificial system, is modified locally, between nearby inputs, as a function of the temporal relation of incoming signals and as a function of the repetition of those temporally related signals. In our computer-based neural network, the synaptic weights would not be modified as a function of the output signals from a common target cell. The input to an element in our network would have to be enhanced only as a function of the timing of that input in relation to other inputs to that element and not as a function of the element's own firing.

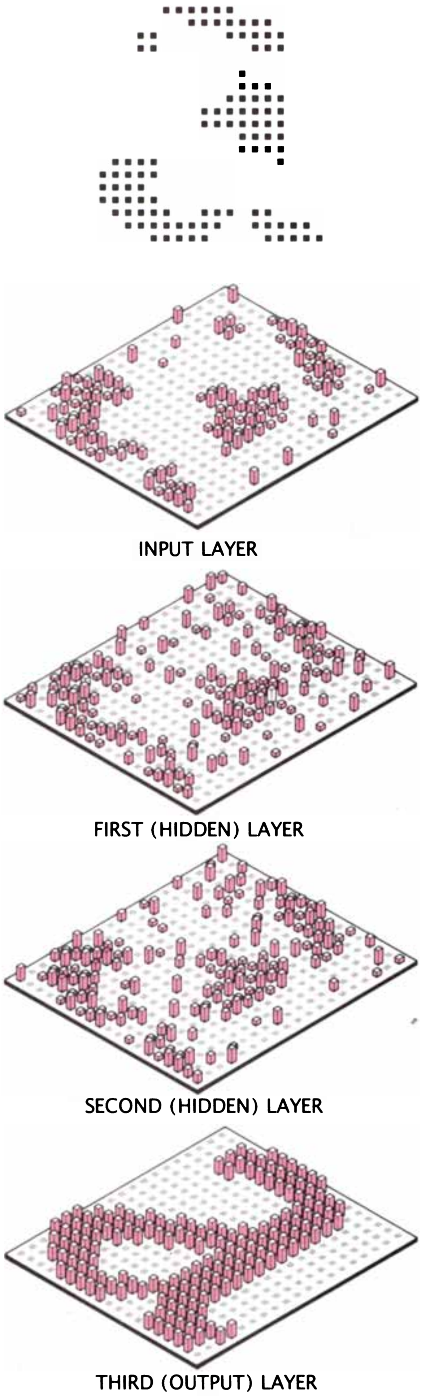
Another design principle was suggested by reflection on what the memories of human beings achieve. When a person stores a pattern (a face, a name, a melody and so on), he or she matches a representation of that pattern with the patterns perceived in the environment. A theoretical memory system should have the same capability, that is, it should have not only some means of storing a pattern for later recall but also a way to represent an image or pattern in real time, just as areas of the brain represent events in human experience as they occur in real time.

An analogy with Pavlovian conditioning may help define the real-time and recall representation functions. An unconditioned stimulus such as the smell of meat, rotation or a puff of air will elicit real-time electrical responses within neural systems and a stereotypical behavioral response. The unconditioned input signal elicits reproducible electrical and ultimately behavioral outputs. Nothing is remembered; the input information simply "flows through" the neural system by means of pathways that are genetically determined or "hard-wired."

A conditioned stimulus such as the ringing of a bell, a flash of light or a tone also elicits real-time electrical responses, but they are not the stereotypical electrical and behavioral responses elicited by the unconditioned stimuli. As a result of the temporal association of the conditioned stimulus with the unconditioned stimulus, the conditioned stimulus elicits new electrical and behavioral responses. The input information from the conditioned stimulus flows along a new pathway, one that is not hard-wired, not genetically dictated but formed in the course of the learning experience. I describe this new pathway as a "collateral" pathway: a pathway whose functional efficacy is determined by the learning experience.

The distinction made between flow-through and collateral pathways suggests another principle of design for a theoretical system of memory. Flow-through pathways should be provided for real-time representation of patterns, and collateral pathways should likewise be provided for remembered or recalled representations. The design should specify that the weight of interaction between elements in the flow-through pathways be constant and high (so as to allow for efficient information transfer) and that the weight of interaction between elements in the collateral pathways be initially minimal (so as to allow no effective information transfer) and

© 1989 SCIENTIFIC AMERICAN, INC



PATTERN RECOGNITION by the author's artificial network operates according to many of the same rules demonstrated by biological systems. When a network is trained to recognize a pattern, such as the lowercase "a" above, the receiving sites participating in the recognition are given more "weight" than those that are not participating—that is, their excitability is enhanced. Here synaptic weight is represented by the elevation of the elements in the layers. Enhancement helps to link together the neurons involved in a recollection when only a piece of a pattern is presented. Thomas P. Vogl of the Environmental Research Institute of Michigan helped to design this drawing.

modifiable as a function of temporally related inputs.

Other features of our theoretical design arise from past observations of biological neural systems. For example, successive aggregates of neurons contain neurons that receive input from large numbers of neurons in preceding aggregates. A theoretical network should also have layers and should also show convergence of signal transfer from the elements of one layer to the elements of the next. Finally, many studies, including those in the horseshoe crab *Limulus* and in *Hermissenda,* demonstrate that synaptic inhibition enhances the contrast between received signals. Borders or edges between fields of visual stimuli, for example, are made sharper by inhibition between neighboring neurons. Such inhibition between elements at a certain spatial distance from one another should be incorporated in a theoretical network as well.

Using these features and others derived from biological systems, my colleagues and I developed a computer implementation that we called DYSTAL (for DYnamically STable Associative Learning). We then tested its ability to "learn" patterns: to store patterns and recognize them later on. Our system successfully acquired patterns such as alphabetic letters and letter sequences and recognized them even when only parts of the patterns were subsequently presented, very much the way a person recognizes a famous face suggested by the sparse strokes of a cartoonist.

The network was truly learning the pattern, since the relation of input to output was in no way preprogrammed within the network. In contrast, many artificial networks that are not based on biological architecture acquire a pattern by reducing errors in comparison to a preprogrammed standard sometimes called a "teacher"; no new information is acquired by the system as a whole. The number of presentations DYSTAL needs to acquire a pattern is quite similar to the number of presentations necessary to train *Hermissenda* or the rabbit; networks that are not based on biological systems usually require thousands or even tens of thousands of presentations to acquire a pattern.

Perhaps the most important characteristic unique to DYSTAL is its ability to accommodate increasing numbers of elements without demanding prohibitive amounts of computer power. In many artificial networks every element is connected to every other element, so that when the number of elements is increased, the number of interactions between the elements rises exponentially. That makes it almost impossible for existing computers to perform the required processing in models having more than 100 elements. Even if the connections between elements in such networks were to remain constant as the number of elements increased, the computation time for the network would not increase linearly because of the way such networks arrive at equilibrium.

Nonbiological artificial networks attain equilibrium by an iterative adjustment of the weight of all the connections so that each weight conforms to a fixed internal standard. It is inherent in such processes that they require more iterations per element as the number of connections increases. In DYSTAL, however, the weights of the connections are not compared with a fixed value; rather, they arrive at a dynamic equilibrium in which the increases and decreases of weight over a set of pattern presentations are equal and no net weight change takes place. As with permanent memory storage, the weight changes can become irreversible when a certain threshold weight level is exceeded. Thus, DYSTAL's computations require less computer power than iterative, nonbiological networks do.

Equations have been derived for the cumulative effect of weight increases and decreases associated with well-defined input patterns. The implications of this derivation go beyond the potential to precompute network weights: they make it possible, for the first time, to compute the internal representations by means of which artificial networks store memories. The equations may one day help neurologists to compare the internal representations of artificial and biological networks. Such comparison would significantly advance the understanding of biological systems and the design of artificial ones.

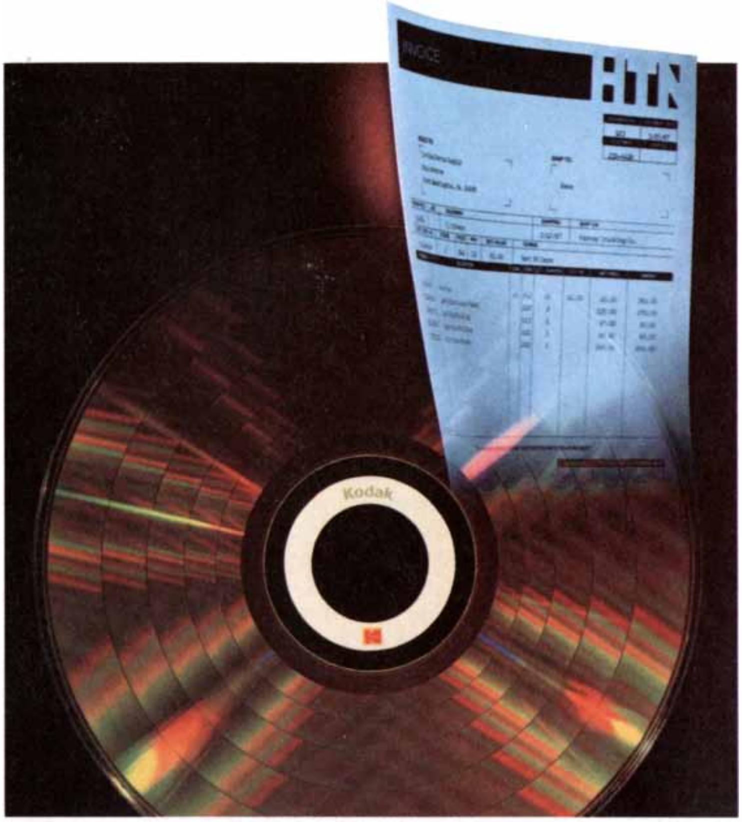
FURTHER READING

CALCIUM-MEDIATED REDUCTION OF IONIC CURRENTS: A BIOPHYSICAL MEMORY TRACE. Daniel L. Alkon in *Science*, Vol. 226, No. 4678, pages 1037-1045; November 30, 1984.

A SPATIAL TEMPORAL MODEL OF CELL ACTIVATION. Daniel L. Alkon and Howard Rasmussen in *Science*, Vol. 239, No. 4843, pages 998-1005; February 26, 1988.

MEMORY TRACES IN THE BRAIN. Daniel L. Alkon. Cambridge University Press, 1988.

© 1989 SCIENTIFIC AMERICAN, INC



The KODAK Optical Disk System 6800...for people with a lot to store, and not much time to process it. The system has the highest capacity and the highest transfer and access rates.

If your business needs to store and retrieve significant amounts of image-intensive or alpha-numeric data, specify the KODAK Optical Disk System 6800. It offers unparalleled storage capacity, flexibility and speed, in the smallest footprint per megabyte available today.

The KODAK Automated Disk Library can put you on-line to over one terabyte—that's 1000 gigabytes —of stored data in just 6.5 seconds. Available in several configurations, the library requires as little as 13 square feet of floor space.

Additionally, the KODAK Optical Storage Interface can connect the system to a wide range of minicomputers or IBM® mainframe host environments.

For a complete package of information, call 1 800 445-6325, Ext. 993A.

SO MUCH DATA, SO LITTLE TIME. CALL KODAK

The new vision of Kodak

Kodak

© 1989 SCIENTIFIC AMERICAN, INC

Spin Glasses

Their traits arise from disorderly, discordant magnetic interactions among atoms. Mathematical models of spin glasses are prototypes for complex problems in computer science, neurology and evolution

by Daniel L. Stein

Dirt can be brushed aside or swept under the rug, but sooner or later it demands attention. In the physical sciences "dirt" can be disorder in structures, impurities in materials or conflicts among interactions. Dirt spoils order. Sufficient amounts of randomness, imperfection and discord may destroy the intrinsic symmetries that simplify physical description tremendously. For most of the history of physics, dirt was brushed aside, and instead physicists investigated orderly systems, such as perfect crystals. By the early 1970's, however, they felt compelled to confront disorder; the dirt that had accumulated in the corridors of science began to loosen. Initially, a little disorder was added to a perfect crystal, for example, to gain an understanding of glasses, where atoms are frozen in random positions in space. Such efforts failed. Excluding too much disorder from inherently dirty systems is like attempting to study a clean mud puddle.

One of the most successful attempts to understand disorderly systems has been the study of spin glasses. The composition of a spin glass is unremarkable—perhaps a few iron atoms scattered in a lattice of copper atoms—but its magnetic properties are confoundedly complicated and sometimes tantalizingly unpredictable. "Spin" is the quantum-mechanical spin from which magnetic effects arise; "glass" refers to the disorder in the orientations and interactions of the spins. The spin glass is the prototypical dirty system. The techniques developed for investigating spin glasses have been applied to the study of complex problems in fields as diverse as computer science, neurology, biochemistry and evolution.

DANIEL L. STEIN is associate professor of physics at the University of Arizona. He was educated at Brown University and Princeton University, where he received his Ph.D. in 1979. He was the first director of the Summer School on Complex Systems in Sante Fe, N. M., in 1988. His work includes computational complexity, spin glasses, structural glasses and biophysics.

The intriguing traits of spin glasses, their dynamics, their complexity—all stem from magnetic interactions among their atoms. Certain atoms can behave as if they were tiny bar magnets. They generate and are subject to magnetic fields. The direction and magnitude of magnetic effects can be described in terms of a vector quantity known as the magnetic moment. If a bulk material whose atoms behave magnetically is exposed to an external field, the magnetic moments will try to align in a particular direction. In some materials this alignment of magnetic moments can also occur as a result of strong, internal effects associated with atomic structure.

One effect aligns the magnetic moments in the same direction. This alignment accounts for iron's strong magnetic properties and hence is called ferromagnetism, although it is found in cobalt, nickel and many other materials as well. Ferromagnetism results from the quantum-mechanical nature of the inner electrons of these metals, which makes it energetically favorable for the magnetic moments of adjacent atoms to be parallel.

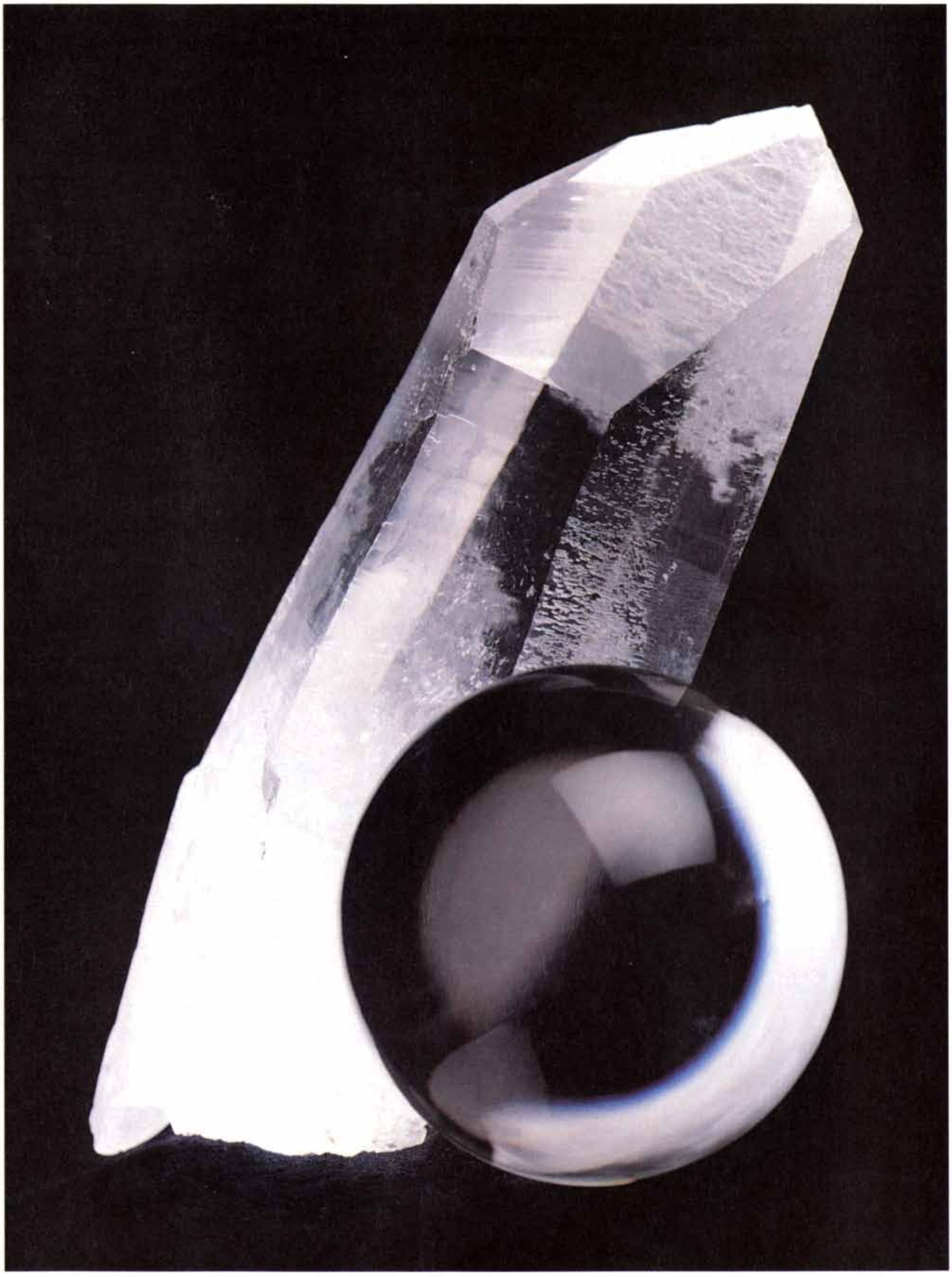
In other words, if the magnetic moments of two adjacent ferromagnetic atoms point in the same direction, energy must be added in order to flip one of the magnetic moments to the opposite direction. Conversely, if the moments point in opposite directions, energy is released when the moments are aligned. Hence, the total magnetic energy is minimum if the magnetic moments of all the atoms point in the same direction.

Adding thermal energy to a ferromagnetic material can affect the alignment of spins. If pure iron is heated to high temperatures, the thermal energy overwhelms the ferromagnetic interactions, so that the direction of each magnetic moment changes randomly from instant to instant. A snapshot of the iron atoms would reveal that on average as many magnetic moments point up as down, left as right, forward as back. The vector sum of all the magnetic moments, or the net magnetization, is zero. (To be precise, the probability is overwhelming that the net magnetization is exceedingly tiny.) In this phase the iron is known as a paramagnet.

As the temperature of iron is lowered, interactions between magnetic moments become dominant, and the moments seek to align themselves in a lower-energy state. At a critical temperature of 771 degrees Celsius, the arrangement of atoms changes suddenly and radically, so that most of the moments become aligned in the same direction. (The magnetization appears to be absent in an ordinary iron sample because the iron undergoes another complicated process that breaks the ordered state into domains. In each domain, however, the magnetic moments all point in the same direction.) At the critical temperature, the iron has undergone a phase transition from paramagnet to ferromagnet.

In other types of materials, in contrast, a different kind of order prevails in their low-energy states. Neighboring atoms of chromium, for example, seek to align their magnetic moments in opposite directions; if the moment of one atom points up, the moment of an adjacent atom points down. Because this is the opposite of iron's behavior, the property is known as antiferromagnetism. Like a ferromagnetic material, chromium has a critical temperature at which it changes from being a paramagnet (random arrangements of magnetic moments) to being an antiferromagnet (opposite

© 1989 SCIENTIFIC AMERICAN, INC



CRYSTAL AND GLASS are quite similar in appearance but are distinct structural phases of matter. A crystal is a solid, but a glass is actually a slow-flowing liquid. Similarly, spin glasses may embody a permanent orientation of the poles of magnetic atoms and constitute a unique phase. Alternatively, they may represent a very slow-changing orientation of atoms.

© 1989 SCIENTIFIC AMERICAN, INC

arrangement of magnetic moments).

Spin glasses, remarkably, display both ferromagnetic and antiferromagnetic properties. In one kind of spin glass, a nonmagnetic host metal is diluted with a small number of atoms that have a nonzero magnetic moment, or spin. Two adjacent magnetic atoms in such a dilute magnetic alloy can interact either ferromagnetically or antiferromagnetically. For example, if a few parts of iron are mixed with, say, 100 parts of copper, the iron atoms, which usually interact ferromagnetically, can now also interact antiferromagnetically. Although the reason is rooted in the subtleties of quantum theory, the phenomenon can be described qualitatively.

Each of the conduction electrons that travel freely through copper has a spin, which is influenced by an iron atom in a rather strange way. At a given distance, the atom aligns the spin of the conduction electrons parallel to its own spin; a little farther away the spin of the conduction electrons is antiparallel to the iron's spin. Still farther away the spin is parallel, and so on. Thus, the iron atom is at the center of a succession of concentric spheres of decreasing influence in which its effect is alternately ferromagnetic and antiferromagnetic.

Because the conduction electrons of the host metal mediate the interaction between two neighboring magnetic atoms, the interaction may be either ferromagnetic or antiferromagnetic, depending on the distance between the atoms. Therefore, in a spin glass composed of many atoms of one metal dispersed in a matrix of another metal, roughly half of all pairs of atoms will interact ferromagnetically; the other half will interact antiferromagnetically. In half of the cases, therefore, the magnetic energy of a pair of iron atoms is lowered if their spins are parallel; in the other cases the energy is lowered if their spins are antiparallel.

The result of this dual behavior is that an atom of a given spin may not be able to orient itself to satisfy its interaction with all the other atoms in the spin glass. Imagine three iron atoms distributed randomly through a copper lattice. The first atom interacts antiferromagnetically with the second, whereas the interactions between the first and third and between the second and third are ferromagnetic. There is no way to satisfy all the interactions at once. If the spin of the first atom points up, for example, then the spin of the second must point down. Yet the third is supposed to align its spin in the same direction as both the first (spin up) and the second (spin down). Any arrangement will violate at least one of the interactions. A system whose interactions cannot all be satisfied simultaneously is "frustrated."

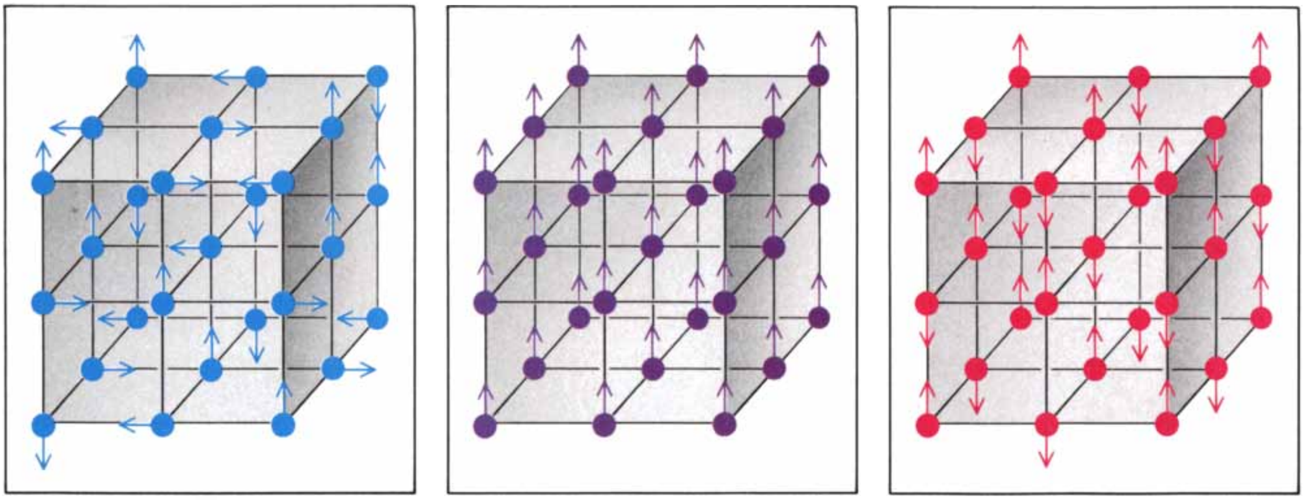
One direct result of frustration is that a spin glass can have many low-energy states. For instance, if the first, second and third iron atoms are respectively either up, down, up or up, down, down, they are in their lowest-energy state because the number of violations is as low as it can be.

The implications of frustration stretch far beyond spin-glass physics to complex problems in many other fields. In the case of spin glasses, the lack of a single low-energy state is related to the question: is the spin glass a new phase of matter, or is it just an extremely sluggish paramagnet? A change from liquid to crystal or paramagnet to ferromagnet as the temperature falls is a true phase transition: the resulting states retain a distinct order for as long as the temperature is maintained. On the other hand, even though common glass seems to represent a new phase, it is fundamentally a liquid: it flows, albeit at such a staggeringly slow rate that it seems to be a solid.

The spin glass could be a distinct phase of matter, whose magnetic order, or alignment of spins, will persist for as long as low temperatures are maintained. On the other hand, they could be paramagnets whose dynamic properties have slowed down so much that they only appear to constitute a static phase. If the spins of a spin glass held at a low temperature were observed to change orientation, one could conclude that it was merely a paramagnet. For such a definitive test, however, the spin glass would probably have to be watched for a period greater than the age of the universe.

One thing that can be done in the laboratory is to look for evidence of a phase transition, a sudden change in the magnetic and thermodynamic properties of the spin glass at some critical temperature. Unfortunately, measurements of different properties have provided conflicting answers.

In 1970 Vincent D. Cannella, John A. Mydosh and Joseph I. Budnick of Fordham University studied the magnetic properties of iron-gold alloys. They measured among other things the



PARAMAGNET, FERROMAGNET AND ANTIFERROMAGNET are diagramed. Atoms behave as if they were tiny bar magnets. Each arrow indicates an atom's "north" pole. The atoms of a paramagnet (*left*) move, casting their poles randomly. In a ferromagnet (*middle*) all the poles are aligned in the same direction. In an antiferromagnet (*right*) adjacent atoms align their poles in opposite directions. A spin glass can resemble a frozen paramagnet: a permanent, random orientation of poles.

© 1989 SCIENTIFIC AMERICAN, INC

alloy's magnetic susceptibility, the change in the magnetization of a material as an external magnetic field is changed in small increments. They, along with many later workers, found an abrupt change in the susceptibility at a critical temperature as the magnetic field vanishes. This sudden change indicated a phase transition.

Other experiments have pointed to the opposite conclusion. One would expect a phase transition to be distinguished by an abrupt change in a thermodynamic quantity such as the specific heat: the change in the amount of heat necessary to raise the temperature of a material by a given amount. No such abrupt change has been observed. Rather, the specific heat of many different types of spin glasses shows a broad, continuous curve with a maximum at a temperature that is typically some 20 percent higher than the transition temperature indicated by the susceptibility measurements.

In addition, some evidence suggests that measurements of magnetic and thermal properties are conducted over too short a time to allow the spin glass to react fully to changing conditions. Whether a phase transition has been seen in laboratory spin glasses is therefore still in dispute.

Whereas contradictory results plague experimentalists, theorists are hampered by their unfamiliarity with thermodynamics of disordered systems. During the past decade, the bulk of theoretical work has been done, for reasons of simplicity and tractability, on patently unrealistic models of spin glasses. Even so, the effort to understand such simple models has proved to be a major theoretical achievement and has uncovered many surprising characteristics.

Sam F. Edwards of Cambridge University and Philip W. Anderson of Princeton University introduced a much studied "short range" spin-glass model in 1975. In their model, spins were positioned at the vertexes of a cubic lattice. A spin interacted only with its nearest neighbors, but a given interaction had an equal chance of being ferromagnetic or antiferromagnetic. Edwards and Anderson assigned numerical values to the interactions at random. (The numbers indicated the strength and type of the interactions: positive numbers for ferromagnetic interactions, negative numbers for antiferromagnetic interactions.)

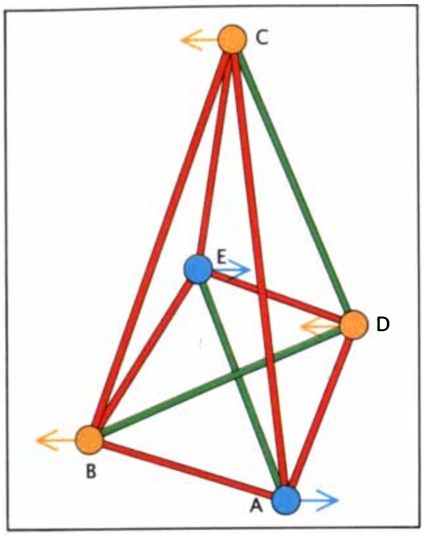
Now imagine tracing a loop, beginning with one interaction between a spin and its neighbor, moving to an adjacent interaction and continuing until you return to the original atom. If you record the values of the interactions around the loop and multiply them, the property of frustration holds whenever the product is negative. Because a typical lattice contains many frustrated loops, it is extremely difficult to determine, for all the sites, the spin values that minimize the total energy of the system. In fact the system may have many low-energy states, which may have no relation among them. It soon became apparent that existing mathematical methods could not reveal low-energy behavior for even this relatively simple model.

Soon after Edwards and Anderson's work, David Sherrington of Imperial College in London and Scott Kirkpatrick of the IBM Thomas J. Watson Research Center proposed an "infinite range" model. They assumed that on the average each spin interacted equally strongly with every other spin in the system. They hoped their assumption, while much less realistic, would yield a readily solvable model. The model did display a true phase transition from a paramagnetic phase to a spin-glass phase, in which spins were frozen in some permanent random arrangement. Attempts to sort out the properties of the spin-glass phase, however, turned out to be much more difficult than expected.

In 1979 Giorgio Parisi of the University of Rome derived a solution for the Sherrington-Kirkpatrick model. The solution is still unproved, although evidence for it is good. At the time of its proposal, however, it was so arcane and different from previous descriptions of phases of matter that four years passed before its properties were understood in a physical sense.

The Parisi solution demonstrates that, below the critical temperature, the Sherrington-Kirkpatrick spin glass can be frozen into one of many possible phases, each representing a low-energy state. The relation between these states, however, is not a simple symmetry transformation, such as inverting every spin. To move from one low-energy state to another, a significant fraction of all the spins must be flipped. In 1977 the work of Anderson, David J. Thouless of the University of Birmingham and Richard G. Palmer of Princeton reinforced the idea that many nontrivial solutions could exist in a large-scale model. This striking feature of the Sherrington-Kirkpatrick model has important implications for problems in computer science, biology and other fields, as I shall discuss.

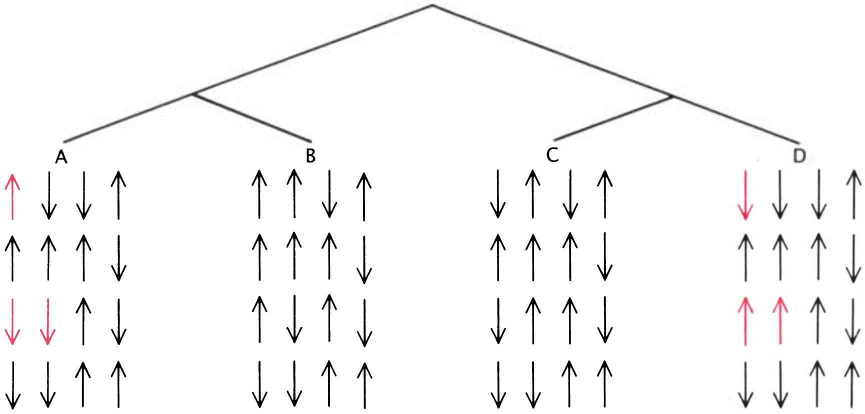
Marc Mézard, Nicolas Sourlas and Gérard Toulouse of the École Normale Supérieure, Miguel A. Virasoro of the Marconi Institute in Rome and Parisi showed in 1984 that these low-energy states nonetheless do display certain relations. In particular, the states can be arranged in a hierarchical manner, as in a family tree or an evolutionary diagram, based on the distances between them. To determine the distance between states, a map giving every spin's orientation is considered for each state. The maps are superposed in pairs so that a spin from one state can be compared with a spin in the corresponding location of the other state. The spin differences between the two states are added up to yield the distance. Low-energy states tend to cluster together in the tree: "siblings" differ the least in distance, then "first cousins" and so on. The emergence of a hierarchical structure in this manner is very unusual for physical systems. The structure is reminiscent of biological systems and other systems that evolve in some sense and may be a feature common to many disordered systems.



FRUSTRATION develops in a spin glass when magnetic interactions conflict. In the arrangement shown, the green and red lines represent interactions that align the poles in the same direction and in the opposite direction, respectively. The arrows represent the poles of one arrangement that satisfies all the interactions except the one between *B* and *C*. In this case no arrangement of the poles can satisfy all the interactions.

Yet the recent studies of more realistic short-range models by the late William L. McMillan of the University of Illinois at Urbana-Champaign, Daniel S. Fisher and David A. Huse of the AT&T Bell Laboratories and Alan J. Bray and Michael A. Moore of the Uni-

© 1989 SCIENTIFIC AMERICAN, INC



SPIN-GLASS "TREE" describes the relation between low-energy states of the Parisi solution. Superposing two of the states and counting the spins that point in opposite directions yield the distance between states. The colored arrows show that the distance between states *A* and *D* is three. The first level of the tree connects states separated by a distance of two; the second level connects states separated by three.

versity of Manchester seem to indicate that the hierarchical structure is a pathological feature of the infinite-range Sherrington-Kirkpatrick model. Although the validity of these new theories is still debated, they reproduce many of the properties observed in real spin glasses. These theories predict a low-temperature spin-glass phase consisting of only two lowest-energy states—not at all like that of the Sherrington-Kirkpatrick model. Explaining how short and infinite-range spin glasses relate to real ones and to each other remains a major challenge for theoretical physicists.

In spite of relative ignorance about the nature of spin glasses, some bold, exciting and perhaps even foolhardy efforts have been made during the past decade to apply what is now understood to some outstanding problems in computer science, neurology and biology. What features could spin glasses possibly share with these problems? Many of them cannot be simplified to problems that have only a few variables. Instead, like spin glasses, they involve a huge number of variables that interact in a nonuniform manner. Typically, just as with frustration in spin glasses, the constraints on the problems cannot all be satisfied simultaneously. The problems therefore often have many possible and apparently unrelated solutions.

Spin glasses have provided a mathematical framework that possesses all these interesting properties and more. The theory of spin glasses was therefore a natural starting point for rudimentary models of these other complex systems. This theory has made it possible to simulate certain properties of systems that hitherto had been difficult or impossible to model.

Some of the first applications of spin-glass mathematics took the form of computer algorithms for solving combinatorial-optimization problems. A well-known example is the traveling-salesman problem: What is the shortest distance a salesperson must travel to visit a given number of cities and return to the city of origin? Finding the answer may seem to be straightforward: calculate the distance between each pair of cities, add up the distances for every possible combination of pairs that completes the journey and choose the shortest combination. Although this strategy works for a few cities, trouble lurks beyond. As the number of cities increases, the number of possibilities explodes.

If the number of cities is five, say, a computer could easily calculate the 12 possibilities. With 10 cities, a computer could still handle the 181,440 possibilities. Yet for just 25 cities the number of possible journeys is so immense that a computer evaluating a million possibilities per second would take 9.8 billion years, almost two thirds the age of the universe, to search through them all.

Usually, combinatorial-optimization problems, such as this one, involve a large number of variables and constraints, an enormous number of possible combinations and a "cost" function, which describes all the possible values of the quantity to be optimized. (In the traveling-salesman problem, the cost function is the distance of the journey as a function of each possible route.) Simply stated, combinatorial-optimization problems ask: what is the lowest-cost solution?

For many of these problems, even those that have a large number of variables, extremely clever algorithms can find the lowest-cost, or "globally" optimal, solution in a relatively short time. For certain types of combinatorial-optimization problems, however, it is strongly believed that an algorithm cannot be devised to find the best solution in a reasonable amount of time for every instance of the problem. These "hard" problems often have practical applications in logic, robotics, language theory and data storage and retrieval. It may come as no surprise that one of these hard problems is finding the lowest-energy arrangement of spins that satisfies the three-dimensional Edwards-Anderson model of a spin glass.

An alternative strategy for difficult combinatorial-optimization problems is to search for "locally" optimal solutions. These relatively low-cost solutions cannot be improved by any small rearrangement of the assigned values, such as flipping a few spins in a spin glass or rearranging the order in which a handful of cities are visited in the traveling-salesman problem. Although one might be reluctant to settle for a "good" locally optimal solution, the alternative may be to wait until sometime in the next century for a computer to calculate the best solution by brute force.

For those not willing to wait around that long, Kirkpatrick, Charles D. Gelatt, Jr., and Mario Vecchi of the IBM research center invented a computational tool. They designed a computer algorithm, called simulated annealing, that finds locally optimal solutions in reasonably short times. Physical methods for reaching a low-energy state in a spin glass form the basis for the algorithm.

The energy of a specific spin glass can be described as a function of the system's state—a particular arrangement of spins. If each spin can point in one of two directions (up or down, say), the total number of distinct states is two to the power of the number of spins in the system. The system defines whether any pair of spins will interact ferromagnetically or antiferromagnetically, and the energy of a particular state depends on how many of the spins in that state satisfy the interactions. If they are all satisfied, the state has minimum energy; if not, it has some higher amount.

The energy as a function of all the states describes a surface in a space whose dimensions equal the number

© 1989 SCIENTIFIC AMERICAN, INC

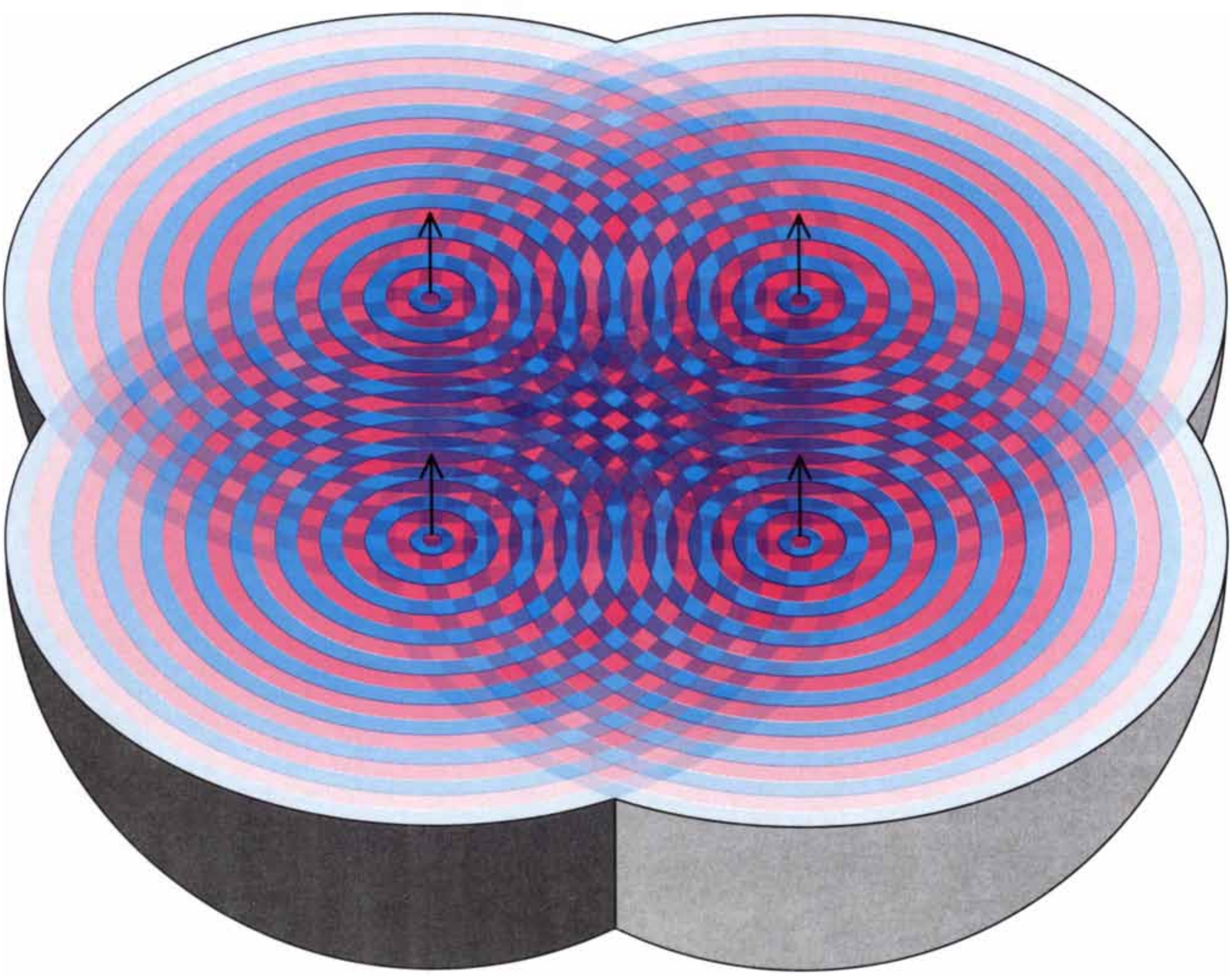
of spins. Instead of trying to visualize such a geometry, imagine the function as a mountain range: the height at each point represents the energy of each state. Suppose the system is currently in a particular high-energy state: it is an athlete on a mountain bicycle on one of the peaks. The athlete's quest is to find the deepest valley—the state of lowest energy—without a map.

The athlete coasts down the mountainside until she reaches a nearby valley. Skeptical that luck would bring her into the deepest valley on the first try, the cyclist pedals out of the valley, reaches a mountain pass and coasts down into an even deeper valley. After pedaling out of many valleys and over many mountains, the athlete believes she has found the deepest valley, although she cannot be certain, since many valleys remain unexplored.

To search for a low-energy state of a spin glass requires heating and cooling—annealing—analogous to pedaling and coasting. If the temperature is extremely low, the system will stay in even a shallow valley for a very long time. As the temperature is raised, the system has more energy available to go exploring, so to speak. Its spins can readily flip around; hence, it has a higher probability of escaping the shallower valleys and can try many more possible spin configurations, some of which may have lower energy than the starting state.

A simple algorithm for finding a relatively low-energy state of a spin glass, then, is to simulate a high temperature (when the system can in principle try any state) and slowly cool the system so that it settles into a state of lower energy. If, at an early stage, it temporarily gets stuck in a high-lying valley, it still has a good chance of escaping over the nearest pass to look for a deeper (lower-energy) valley. After many cycles of heating and cooling, the algorithm has a high probability of yielding a good solution—that is, a low-energy state—even though the chance of accidentally finding the best solution in that huge space is extremely small.

For many combinatorial-optimization problems the cost function resembles a rough landscape in a state space, much like the energy function of a spin glass. Simulated-annealing algorithms tackle combinatorial-optimization problems as if the goal were to find a low-energy solution for a spin glass. The cost function becomes energy as a function of state. Although the temperature in these problems has no physical meaning, it can be raised or lowered, like physical tem-



SPIN GLASS can result when atoms interact with conduction electrons. Each arrow points in the direction of an atom's "north" pole. Each atom is at the center of a succession of concentric spheres of decreasing influence (*color*) in which the poles of electrons surrounding the atom are aligned alternately south (*red*) and north (*blue*). The electrons mediate the interactions among the atoms, whose poles can flip because of the influence of other atoms and the surrounding electrons.

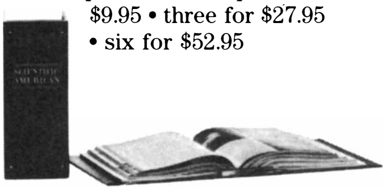
© 1989 SCIENTIFIC AMERICAN, INC

To preserve your copies of SCIENTIFIC AMERICAN

A choice of handsome and durable library files or binders. Both styles bound in dark green library fabric stamped in gold leaf.

Files Each file holds 12 issues.
Price per file $7.95
• three for $21.95
• six for $39.95

Binders Each holds 12 issues.
Issues open flat. Price per binder
$9.95 • three for $27.95
• six for $52.95



(Add $1.00 per unit for postage and handling in the U.S. Add $2.50 per unit outside the U.S.)

Major credit cards accepted for $15.00 or more. Provide card number, expiration date, signature or call Toll Free: 1-800-972-5858

To: Jesse Jones Industries
Dept. SA, 499 East Erie Ave.
Philadelphia, PA 19134

Send me ________
SCIENTIFIC AMERICAN
☐ Files ☐ Binders
For issues dated ☐ through 1982
☐ 1983 or later.
I enclose my check or money order for
$____________ *(U.S. funds only).*

Name ______________________
(please print)

Address ______________________
(No P.O. Boxes please)

City ______________________

State ____________ Zip ________

SATISFACTION GUARANTEED.
Pennsylvania residents add 6% sales tax.

perature, thereby enabling the system to explore regions of state space in search of a good solution. In this way, simulated-annealing algorithms can find low-cost solutions, relatively quickly, for a variety of combinatorial-optimization problems.

In the early 1980's John J. Hopfield of AT&T and the California Institute of Technology proposed another important application of spin-glass mathematics. He realized that a system similar to spin glasses could perform computations and store information, provided it had the appropriate dynamic rules. The system was particularly interesting because it mimicked the architecture of the brain more closely than standard digital computers do [see "Collective Computation in Neuronlike Circuits," by David W. Tank and John J. Hopfield; SCIENTIFIC AMERICAN, December, 1987].

Hopfield's model is made up of simple "neurons," each of which can assume one of two states: on, or firing, and off, or not firing. Whether a neuron remains in its present state or switches depends on the states of all the neurons connected to it. The nature of the computational task determines the pattern of neural connections. Unlike real neurons, those in Hopfield's model interact symmetrically: the effect of one neuron on another is the same as the reciprocal effect.

In some respects this system does resemble the spin glass: it is a collection of variables—of which each can take on two values—that interact in a complex, nonuniform manner. Crucial differences lie in the way that the probability distribution of the interactions of the neural model is chosen and the fact that the interactions may evolve in time, which enables "learning" to take place.

A state-dependent energy function can be defined for this system in the same way as for spin glasses. As before, the result is a rough, mountainous landscape in state space. Energy valleys might correspond to memories to be recalled, patterns to be recognized or other kinds of mental behavior. The neural connections of the system fix the number, location and meaning of the valleys.

A stimulus from the external world determines the initial location of the system in state space—that is, which neurons are firing and which ones are quiescent. A basin of attraction surrounds the lowest point in a valley: the "solution." If the system finds itself anywhere in the basin, it will evolve toward the associated solution. Hence, inputs from the external world initiate the selection of a given solution—recall of a given memory, say.

This type of system differs considerably from linear, sequential algorithms designed for digital computers. The computation employed here, like that controlling the nervous system of an animal, is collective in nature—that is, all parts of the system interact with one another simultaneously, and many of them change their state as the computation proceeds.

In another link between the mathematics of spin glasses and biology, investigators have found an appealing relation to biological evolution. One central question of early evolution is: how does a "soup" of small molecules such as amino acids or nucleic acids evolve into highly organized, information-carrying macromolecules such as proteins or DNA? The exact chemistry of the process is unknown. Yet one can construct a mathematical model of molecular evolution in which an interesting transition from little to much "information" occurs.

Say that the biological information contained in a sequence of monomers (the building blocks of macromolecules, or polymers) is valuable if it helps the polymer to survive. Before a complicated cellular apparatus existed for translating DNA sequences into functional proteins, a polymer's probability of survival may have been directly related to the chemical properties of the actual sequence: ease of replication, folding characteristics, probability of adsorption onto a nearby surface, tendency toward self-entanglement, stability and so on.

In the early 1980's Daniel S. Rokhsar of Princeton, Anderson and I considered two monomers, X and Y, that in principle can represent any two molecules of biological interest. The two monomers are mixed in equal parts. A sequence of events is imposed on the system, building up longer and longer strings of monomers as time passes. The key to generating a large information content lies in a process that exhibits both diversity and selection. Diversity means that a large number of different polymers are possible as an outcome of this sequence of events. If, for example, an X-to-Y bond is much more stable than bonds between X and X or Y and Y, the growth process will always yield polymers with the sequence X, Y, X, Y, X, Y and so on, and nothing is learned from the process. Selection, on the other hand, means that not all polymers have an equal

© 1989 SCIENTIFIC AMERICAN, INC

chance of formation or survival. If a system generates every possible type of polymer, again nothing is gained. The information content of this system can be defined simply by the logarithm of the number of possible polymers divided by the number of existing polymers.

To create both diversity and selection, my colleagues and I studied a rather abstract model in which the interaction between any two monomers in a chain has a roughly equal chance of either increasing or decreasing the sequence's survival probability. For a given chain, the total survival probability per unit of time is the sum of the contributions from every pair. A mathematical model can be constructed that accounts for these factors statistically. The mathematics of such a model is very similar to that of the spin glass.

Like the spin-glass function of state, the survival-probability function has many peaks and valleys in the state space of all polymers in the model. The model demonstrates that not all polymers are equally probable (which seems realistic). Furthermore, as long as the number of "good" polymers increases exponentially with polymer size, sufficient diversity is maintained to generate information.

Whatever else may come of these endeavors, they have helped foster a remarkable cross-fertilization among disciplines as diverse as physics, mathematics, computer science, biology, chemistry and economics. A growing number of investigators are grappling with fundamental problems that embody the uncertainty and disorder inherent in the universe. We have begun to develop a deeper understanding of why these systems are so difficult to treat mathematically, and we may come to a better understanding of these systems in their own terms. We are finally rolling up our sleeves and getting our hands dirty.

FURTHER READING

SPIN GLASS THEORY AND BEYOND. M. Mézard, G. Parisi and M. A. Virasoro. World Scientific Publications, Singapore, 1986.

SPIN GLASSES AND OTHER FRUSTRATED SYSTEMS. D. Chowdhury. World Scientific Publications, Singapore, 1986.

COMPLEX SYSTEMS. Edited by D. L. Stein. Addison-Wesley Publishing Company, in press.

ORIGINS OF ORDER: SELF-ORGANIZATION AND SELECTION IN EVOLUTION. S. A. Kauffman. Oxford University Press, in press.

QUESTAR®

. . . the priceless telescope

Some people say a Questar telescope is expensive; others say it is a priceless possession. We say that its price is a fair one, based on our manufacturing costs, with only a modest profit added. We think it is a price that the serious astronomer is willing to pay for the finest instrument he can own. Not only is he acquiring optical resolution that many claim surpasses theory, but, in addition, will enjoy a mount designed to accomplish all the motions of a great observatory telescope with the utmost ease, on a table top.

Some would say the price of perfection comes high, but we are not sure about that either. In the thirty-eight years we have been making Questars we have seen the price of cars increase sixfold. Questar's price for perfection has slightly more than doubled.

So why not have a Questar telescope? It will give you

- admission to the stars
- a membership card to all the sky spectaculars
- an intimate look into the shadows of the moon
- access to the secret lives of nature's world
- travel opportunities with an easily portable telescope in its own hand luggage
- free special indoor entertainment at ten feet in full color

A Questar is not only a priceless telescope, it is a unique experience.

1988 QUESTAR CORPORATION

QUESTAR, THE WORLD'S FINEST, MOST VERSATILE TELESCOPE IS DESCRIBED IN OUR BOOKLET. SEND $3 TO COVER MAIL COSTS ON THIS CONTINENT. BY AIR TO S. AMERICA, $4; EUROPE & N. AFRICA, $5.00; ELSEWHERE, $5.50. INQUIRE ABOUT OUR EXTENDED PAYMENT PLAN.

QUESTAR

Box 59, Dept. 219, New Hope, PA 18938
Phone (215) 862-5277

© 1989 SCIENTIFIC AMERICAN, INC

Whoever said, "You can't take it with you" didn't have the Sony Video Walkman™ Personal TV/VCR in mind.

It's the first compact portable TV and VCR that can actually be held in the palm of your hand. So if you're someone who's going places, it'll come in very handy.

If you're someone without a lot of time on your hands, there's a built-in 8mm VCR. With an easy-to-use timer. So you can record up to 4 hours of your favorite TV programs. And watch them or your favorite movies in your own good time.

All of which makes the Sony Video Walkman perfect for time-shifting. And place-shifting.

Not only can you use the Video Walkman whenever or wherever you want, you can use it however you want. You can watch its 3" color LCD TV screen with unbelievable sharpness and color fidelity. Or connect it to any other ordinary TV.

But you ain't seen nothin' yet.

The Video Walkman weighs a lightweight 2½ lbs., but it's not light on features. It has high-speed picture search. A sleep timer that turns it off one hour after you've turned yourself off. A worldwide, multi-voltage AC adapter battery charger. And a built-in TV tuner with full VHF and UHF reception.

Which brings us to another feature sure to get a great reception. Our highly adaptable audiovisual input. You'll be sky-high on it because it lets you record and

© 1989 Sony Corporation of America. Sony, Video Walkman, Video 8 and The One and Only are trademarks of Sony. Simulated television picture. All titles used with permission.

© 1989 SCIENTIFIC AMERICAN, INC

play back your Video 8® home movies. Or edit them, no matter what format you shot them on.

Now, if you think all this sounds pretty good wait until you hear the Video Walkman. With its exceptional high-fidelity sound.

Besides getting a lot of enjoyment out of the Video Walkman, you'll get a lot of enjoyment into it. For there are over 1300 movie titles now available on 8mm. Like Crocodile Dundee II. The Accidental Tourist. Coming to America. The Dead Pool. And A Cry In The Dark.

And if all that wasn't enough, the sky's the limit on the home instruction, travel and video magazines available. They're part of the video publishing software on 8mm.

So, if your interests take you out. Or in. Up. Or down. Check out the Sony Video Walkman Personal TV/VCR.

It's the most incredible TV and VCR you'll find anywhere.

SONY®
THE ONE AND ONLY®

© 1989 SCIENTIFIC AMERICAN, INC

Volcanism at Rifts

When the earth's rigid shell is rifted, the ductile rock of the mantle wells up and partially melts. Spectacular volcanic outbursts ensue when the mantle is only slightly hotter than normal

by Robert S. White and Dan P. McKenzie

Sixty-six million years ago the earth rifted open on the west side of India, and huge volumes of molten rock poured onto the land. Close to two million cubic kilometers of lava were released in less than half a million years, blanketing much of west-central India in layers of basalt hundreds of meters thick. The episode may have dealt a major blow to the climatic and ecological stability of the planet, and some theories even blame it for the extinction of the dinosaurs. Yet it was far from unique; many similar cataclysms have taken place on rifting continents during the course of geologic time.

Volcanism at rifts takes another, quieter guise that in the long term amounts to even more volcanic activity. The volcanism occurs steadily but out of sight along the system of mid-ocean ridges that dissects the floor of the world's oceans. These ridges mark rifts at which releases of magma generate some 20 cubic kilometers of new oceanic crust every year.

The contrast between the sudden, catastrophic volcanism recorded in flood basalts and the routine volcanism that generates oceanic crust has led many workers to suggest that the underlying mechanisms must somehow be different. Yet we are now able to explain both kinds of rift volcanism, as well as account for the amount and composition of the magma they produce and for the vertical movements of rifting crust, by a single mechanism—one that is based on simple physical ideas and the familiar processes of plate tectonics.

ROBERT S. WHITE and DAN P. McKENZIE are both at the University of Cambridge. White leads the marine geophysics group and was recently named professor of geophysics; McKenzie holds a personal chair in earth sciences. Both men received a Ph.D. at Cambridge, White in 1977 and McKenzie in 1966. They have spent most of their careers at Cambridge but have also worked extensively at research institutions and universities in the U.S. White and McKenzie have traveled widely to study tectonic, geologic and seismologic phenomena on land and at sea.

The fundamental concept is simple: rift volcanism takes place when tectonic movements stretch and ultimately rupture the earth's outer layer, thereby allowing the underlying mantle to rise into regions of lower pressure, where some components of the hot but solid mantle rock melt. Differences between the catastrophic volcanism at some rifts and the quieter process at others simply reflect differences in the starting conditions—the temperature of the parent mantle in particular. We draw on two kinds of evidence: quantitative descriptions of rock melting and a wide range of observations made on the rifted edges of continents and in the oceans that have opened between them.

From the Laboratory to the Earth

For many years workers have studied the melting of mantle rocks by simulating the temperature and pressure of the deep earth in the laboratory. A cubic millimeter or so of rock that has the same composition as the earth's mantle is raised to a specific temperature and pressure and then rapidly cooled. Any material that has melted freezes into a glass, which can be analyzed to learn the amount and composition of the melt. Increasing the temperature yields more melt and changes its composition, by melting successively more resistant components of the rock. Increasing the pressure reduces the amount of melt by raising the solidus—the temperature at which the rock begins to melt.

In recent years many such experiments have been carried out by, among others, David Green and his students of the University of Tasmania, A. E. Ringwood of the Australian National University, David Walker and Edward M. Stolper of Harvard University, Eiichi Takahaski and Ikuo Kushiro of the University of Tokyo and Toshitsugu Fuji and Christopher M. Scarfe of the University of Alberta. The field has been known for its bitter controversies. Yet the experimental results actually agree very well: one of us (McKenzie), together with Michael J. Bickle of the University of Cambridge, was able to turn the many observations into simple mathematical expressions describing how the amount and composition of the melt depend on the temperature and pressure. Most of the apparent inconsistencies in the experimental results, it turned out, are scarcely larger than the uncertainties in the measurements!

How might these expressions be applied to melting at rift zones? The processes of plate tectonics, including rifting and plate motion, all affect the rigid outer layer of the earth, known as the lithosphere, which includes the crust and the underlying mantle to an average depth of about 120 kilometers. Below that depth the mantle remains solid but becomes ductile—about as ductile as window glass at room temperature, ductile enough to flow on geologic time scales. When the overlying lithosphere is stretched and thinned or ruptured as plates are pulled apart, this ductile mantle, known as the asthenosphere, wells up to take its place. As the mantle rock rises, its pressure falls.

If the upwelling is fairly rapid—a matter of centimeters per year—as it is when plates rift apart, the mantle will not have time to lose heat by conduction and may eventually rise to a level at which its temperature exceeds the solidus. At first only a little melt is produced, consisting of the rock components with the lowest melting temperatures, but at shallower depths the molten fraction grows.

Given the temperature of the upwelling mantle and the amount of

© 1989 SCIENTIFIC AMERICAN, INC

lithospheric thinning, our expressions predict the amount and composition of the melt produced at each level of the rift zone. Summing the results over the entire depth of melting gives the total amount of melt and its net composition. By varying the initial conditions we made a startling discovery: for a particular amount of lithospheric thinning, the magma yield is acutely sensitive to small variations in the temperature of the parent mantle.

How much of this melt actually escapes from the mantle to be released in volcanism or trapped in the crust? The melt is buoyant because it consists of the lighter components of the rock. Whether it can pass through the overlying rock, however, depends on its ability to "wet" the rock grains. A droplet of mercury, for example, is readily trapped on a piece of cloth because it cannot wet the threads. Yet water soaks right through.

Laboratory studies suggest that molten rock behaves more like water than like mercury. Stolper and Walker of Harvard and Donald L. Turcotte of Cornell University have argued that melt should separate rapidly from the residual, unmelted crystals once it amounts to more than 1 or 2 percent of the sample. Detailed calculations by David R. Scott and David J. Stevenson of the California Institute of Technology and by one of us (McKenzie) have confirmed that essentially all the molten rock should escape from the melting zone. The effects of temperature variations on melt production, then, should be fully apparent near the surface.

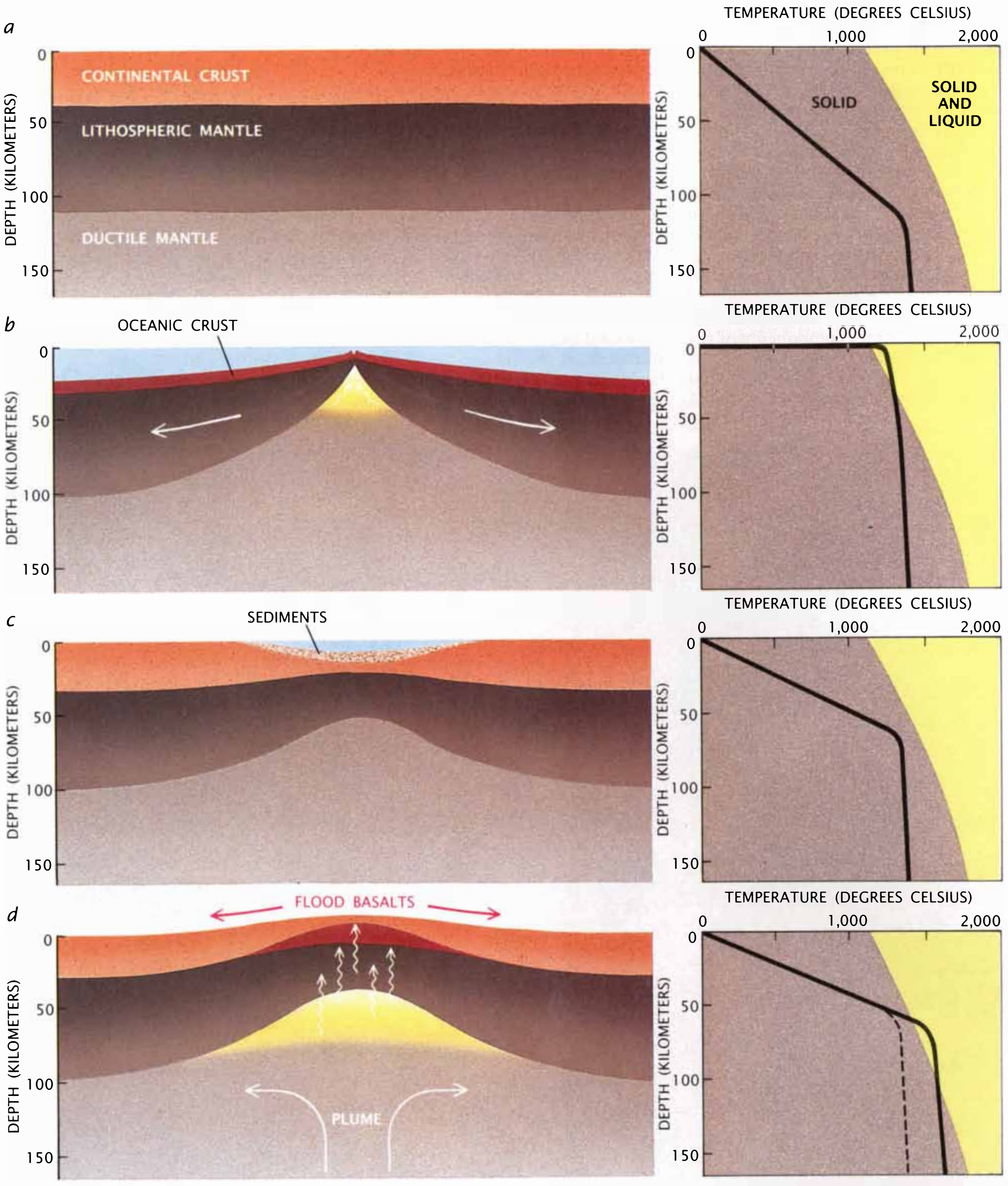
A Natural Laboratory

The earth provides a superb natural laboratory in which to test our model of melt production at rifts: the ocean floor. Each midocean ridge marks a rift between two plates that are spreading steadily apart while volcanic activity continuously adds material to the retreating edges of the plates. Some workers have argued that a local heat source in the mantle accounts for the volcanism, but our model should be able to account for it by decompression melting alone.

At an oceanic rift the lithosphere is being ruptured, allowing the astheno-

STEPPED LANDSCAPE of the Deccan Traps, in west-central India, was formed over a period of about 500,000 years by a series of huge basalt flows; each layer represents a major eruption. The outpouring began some 66 million years ago when a rift opened over a region of unusually hot mantle. The image is by Keith G. Cox of the University of Oxford.

© 1989 SCIENTIFIC AMERICAN, INC



THINNING OF THE LITHOSPHERE (the crust and uppermost mantle) can lead to partial melting of the hot, ductile rock below. In unthinned lithosphere (*a*) the temperature (*thick curve*) increases with depth, but increasing pressure keeps the ductile rock under the lithosphere from melting. (The colors on the graph show the effect of depth on the melting temperature.) Where the lithosphere has been rifted, forming an ocean basin (*b*), the hot mantle rock can well up to the surface. Because the melting temperature falls as pressure is reduced, part of the initially solid material melts as it moves upward. The resulting magma solidifies to form the oceanic crust. If the lithosphere is only stretched, say by a factor of two—as it is in a nascent rift or a sedimentary basin (*c*)—little or no melt is ordinarily produced. The hot mantle rises, and the temperature profile approaches but does not cross the melting curve. Yet if a plume in the mantle has heated the rock by perhaps 200 degrees Celsius (*d*), the same amount of stretching leads to a volcanic outburst. The reduction of pressure in the upwelling mantle generates magma that spills out onto the land as flood basalts and is "underplated" onto the base of the crust.

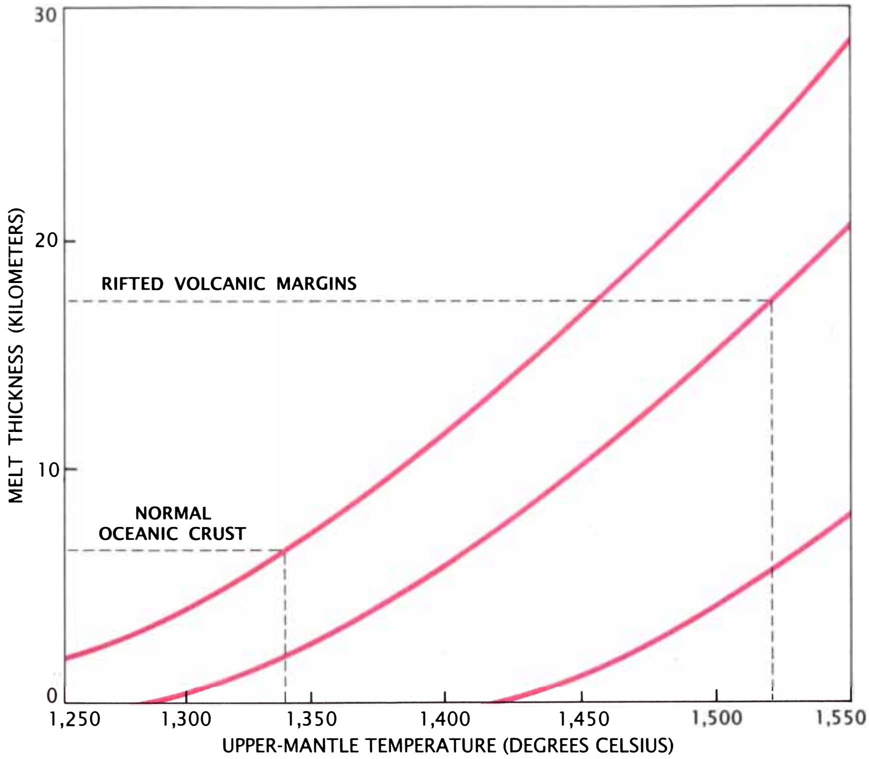
© 1989 SCIENTIFIC AMERICAN, INC

sphere to well up nearly to the surface. The release of pressure generates magma that forms the new oceanic crust. The model's quantitative expressions must be able to account for the thickness of the crust (which indicates the volume of melt) and for its composition.

Air guns or explosions at the sea surface generate low-frequency sound waves that, as seismic waves, can penetrate the oceanic crust and the lithospheric mantle below it. These seismic waves are refracted or reflected back to the surface by differences in rock properties such as seismic velocity and density; it is thereby possible to image the subsurface layers. During the past three decades this technique of seismic profiling has shown that the thickness of oceanic crust is remarkably consistent, averaging between six and seven kilometers.

This consistent thickness is readily explained by our model of melt generation by decompression: the upwelling mantle is decompressed by the same amount regardless of the rate at which the plates move apart. Moreover, given a mantle temperature in the expected range, from 1,320 to 1,360 degrees Celsius, the model predicts the generation of just the right amount of magma. Mantle at 1,340 degrees C should start to melt when it rises to a depth of about 50 kilometers; by the time it reaches the surface an average of 25 percent of the rock should have melted. The result will be oceanic crust of the observed thickness.

Thus, our simple model of decompression melting can account for magma production at midocean ridges and defines a surprisingly precise figure for the temperature of the underlying mantle. Moreover, the oceanic crust's consistent thickness implies that the temperature does not vary by more than 20 degrees over most of the globe. The model also predicts the composition of the melt released at midocean ridges, accurately reproducing the known average composition of oceanic crust. It is encouraging that Emily M. Klein and Charles H. Langmuir of the Lamont-Doherty Geological Observatory of Columbia University have also found evidence, in a survey of global data, that a consistent mantle temperature at the midocean ridges is correlated with consistent crustal thickness and composition.



THICKNESS OF THE MELT LAYER generated at a rift depends on the degree to which the lithosphere is thinned and on the temperature of the underlying mantle. The authors' model predicts that thinning by a factor of two (*bottom curve*) above mantle at its normal temperature of 1,340 degrees C generates no melt; thinning by a factor of five (*middle curve*) generates small amounts. A complete break in the lithosphere (*top curve*) gives a melt layer 6.5 kilometers thick—the observed average thickness of oceanic crust. If the mantle is less than 200 degrees hotter, however, lithosphere thinned by a factor of five yields enough melt to account for the massive deposits of basalt, 15 kilometers and more thick, seen on the rifted edges of some continents.

Rifting Continents

Every ocean basin begins when a continent is stretched, thinned and finally ruptured. In the same way as it does at the midocean ridges, the rifting allows the underlying mantle to well up and melt as the pressure on the rising material is reduced. What can our model say about the magmatism that results when rifting begins on a continent?

In some continental regions the lithosphere has been stretched by a small amount—by a factor of two, say—after which the stretching has stopped. The equations of decompression melting predict that the limited upwelling should produce little if any melt. Yet such regions do illustrate a phenomenon that shapes all continental rifts: crustal subsidence.

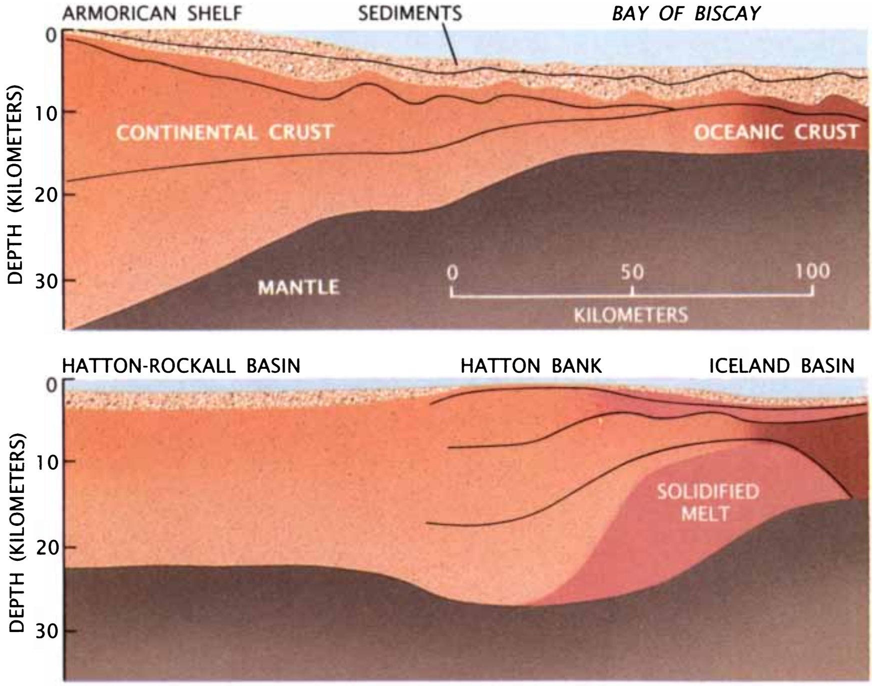
The continents are above sea level because the continental crust floats on the denser mantle below. Lithospheric stretching thins the continental crust, causing it to subside below sea level. Some of the initial subsidence is offset by a second process: the thinning of the underlying lithospheric mantle, which enables hotter (therefore less dense) asthenospheric mantle to take its place. Later the added material cools to the normal temperature of the lithosphere, and further subsidence, called thermal subsidence, takes place.

The result is a basin sunken far below sea level, in which sediments accumulate. The North Sea is a good example: the underlying continental lithosphere has been stretched by about 35 percent, creating a basin in which more than two kilometers of sediment—including the precursors of oil and gas—have accumulated.

When the stretching of the continental lithosphere does not stop, the continent eventually breaks in two, and a new ocean develops between the separating edges. The continental margins that once flanked the young rift record the history of stretching, subsidence and volcanism. Seismic profiling of such margins has shown that the final rupture generally comes when the continental crust has been stretched and thinned beyond a factor of about six. Ordinarily the rift releases little magma until the continent splits apart completely.

At the Biscay margin of France, for example, formed about 120 million

© 1989 SCIENTIFIC AMERICAN, INC



CONTINENTAL MARGINS formed by rifting record differing amounts of volcanic activity. A profile of the European margin at the Bay of Biscay (*top*), based on seismic data gathered jointly by the Institute of Oceanographic Sciences in England, the French Petroleum Institute and the Oceanographic Center of Brittany, shows stretched and thinned continental crust with no evidence of volcanic activity until just before the transition to oceanic crust. In contrast, a profile across Hatton Bank (*bottom*), west of Scotland, reveals layers of igneous rock many kilometers thick both above and below the thinned continental crust. Because the rifting that formed this margin took place above unusually hot mantle, it triggered a volcanic outburst.

years ago by the opening of that part of the Atlantic, the European continent ends in a 200-kilometer-wide wedge in which the crust gradually thins in a seaward direction, beginning at an initial thickness of more than 30 kilometers. The thinning caused the crust to subside, so that it now lies well below sea level. Little evidence of volcanism appears until, far offshore at the edge of the continental shelf, the crust is reduced to less than a quarter of its original thickness. There it is edged by a narrow transition region containing igneous rocks, which in turn gives way to seven-kilometer-thick oceanic crust produced by decompression melting at the new oceanic spreading center.

Volcanic Margins

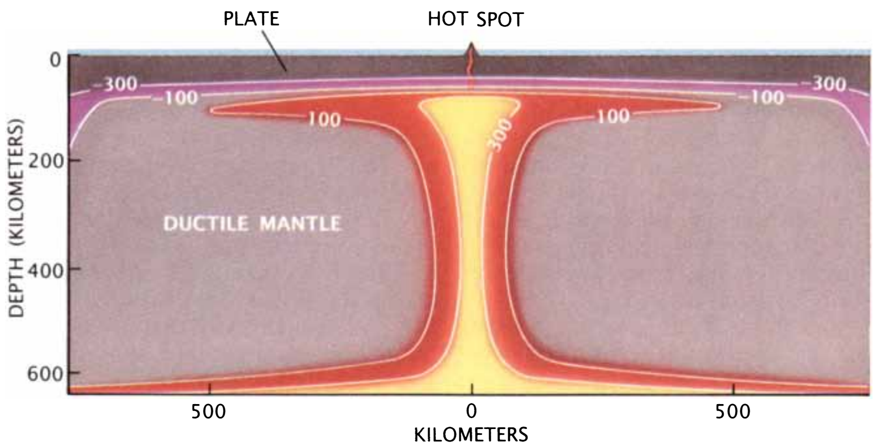
Not every rifted margin shows this orderly transition from continental to oceanic crust. In about 1980 Karl Hinz of the Federal Institute for Geosciences and Natural Resources in Hanover, West Germany, discovered evidence for thick layers of rock under the superficial sediments on some rifted continental margins. Hinz proposed that the layers, which are many kilometers thick, resulted from voluminous lava flows that had been erupted onto the thinned continental crust just at the time of rifting.

Hinz's proposal was based on seismic profiling of continental margins around the world, and it was at first hotly disputed, since the layering in many respects resembles the thick stacks of sediments found in submerged river deltas. The controversy was resolved when the Deep Sea Drilling Project took core samples of the rock from the seabed off Norway. The layers were basalt, similar in many respects to the basalts formed at oceanic spreading centers.

Once their characteristic reflection patterns were known, the volcanic layers were recognized on many continental margins. They are now known to extend for 2,000 kilometers along the east coast of Greenland and, on the other side of the North Atlantic, for a comparable distance along Hatton Bank (off Scotland) and the west coast of Norway. Volcanic margins have also been traced along the east coast of the U.S., on opposing coasts of South America and Africa, and off Antarctica, the west coast of India and the west coast of Australia.

It is now clear that vast amounts of lava flooded onto the continental crust during the early stages of many rifts. What is more, in the course of a seismic survey across Hatton Bank in the summer of 1985, we found that shallow igneous layers record only part of the volcanic outburst. Directly under the continental margin, between the thinned crust and the lithospheric mantle, the profiles revealed a massive prism of igneous rock, up to 18 kilometers thick. Other workers have found comparable volcanic deposits "underplated" onto the Norwegian section of the margin.

Nor is the east side of the North Atlantic rift likely to be unique. Possible seismic evidence of underplating has been reported for other volcanic margins, and even before our study of



PLUME of unusually hot rock rising from deep in the mantle has effects well beyond the hot spot that marks the plume axis. The contour lines on this computer model of the mantle underlying the Cape Verde hot spot trace deviations from the average upper-mantle temperature of 1,340 degrees C. The contours show that as the plume meets the lithosphere it is deflected, resulting in elevated mantle temperatures in a region more than 1,000 kilometers across. Other plumes, such as the one under Iceland, create temperature anomalies as broad as 2,000 kilometers.

© 1989 SCIENTIFIC AMERICAN, INC

Hatton Bank some workers, notably Keith G. Cox of the University of Oxford, had suggested that igneous intrusions at the base of the crust underlie all shallow flood basalts.

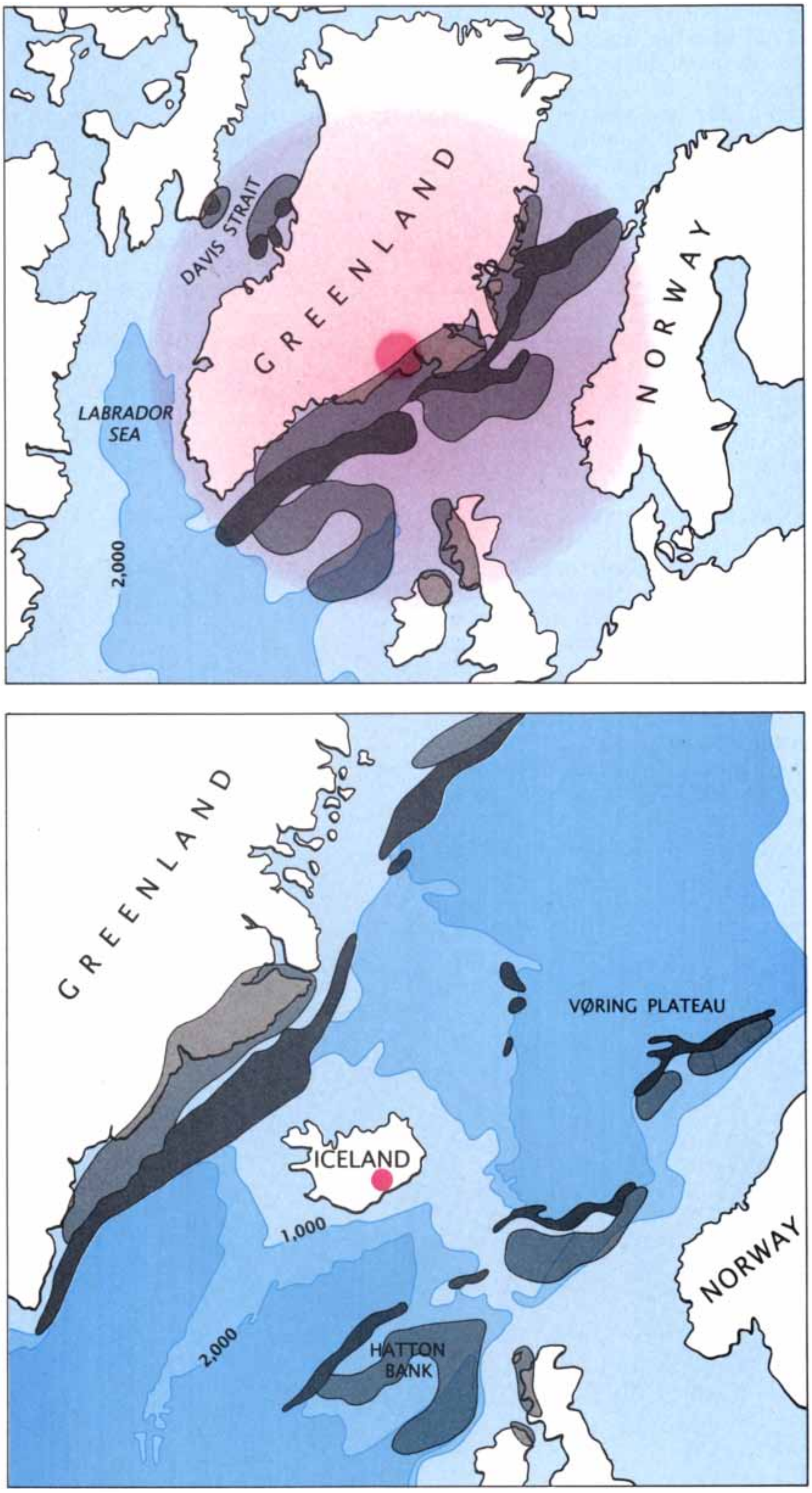
Once the underplated material is taken into account, the volume of magma released along such rifts is seen to be truly enormous. The igneous layers that have been extruded onto the margin of the eastern North Atlantic over a distance of 2,000 kilometers average between four and six kilometers thick and 50 to 100 kilometers wide: a total of between one and two million cubic kilometers of rock. When the rock intruded into the lower crust is added, the total rises to between five and 10 million cubic kilometers, all emplaced within at most a few million years as the North Atlantic rift split the protocontinent.

It is difficult to imagine 10 million cubic kilometers of rock in everyday terms. Picture enough rock to cover the entire U.S., including Alaska, to a depth of about a kilometer. Or think of 20 truckloads of lava being dumped every second, day and night, year after year, for a million years. When other rifted margins, such as the Biscay margin, produced melt so sparingly, how can so huge and sudden an outburst be explained?

The Heat of the Mantle

It was in the tearoom at Madingley Rise, the building at Cambridge where we both work, that we came across the key to understanding these volcanic margins. By that time, early in 1987, McKenzie and Bickle were in the midst of their analysis of laboratory results on melting and had found that the amount of melt produced by mantle decompression rises dramatically with small increases—say by 100 or 200 degrees—in mantle temperature. At about the same time White and Robert C. Courtney (then a graduate student) had finished mapping a large-scale thermal anomaly under the Atlantic Ocean. They had defined an area some 1,500 kilometers in diameter in which the mantle was perhaps 150 degrees warmer than normal. The answer, we realized in our conversation, lay at the intersection of these apparently unrelated lines of work.

The anomaly Courtney and White studied is centered on the Cape Verde hot spot: an isolated area of intense volcanic activity in the equatorial Atlantic that is thought to overlie a plume of unusually hot material rising from deep in the mantle. Traditionally, hot spots have been viewed as rather



NORTHERN NORTH ATLANTIC was the scene of extensive volcanic activity when it rifted open some 57 million years ago (*top*). The outpouring left massive basalt layers (*dark gray*) and regions of abundant volcanic rocks (*light gray*) along the continental margins. The positions of the Iceland hot spot (*red*) and of the surrounding zone of unusually hot mantle are superposed on the ancient geography. The extent of the temperature anomaly is just right to account for the length of the volcanic margins and also for igneous deposits west of Greenland in the Davis Strait. Today a broad region of anomalously shallow seafloor around Iceland reflects the continued presence of the mantle plume (*bottom*). The depth contours are in meters.

© 1989 SCIENTIFIC AMERICAN, INC

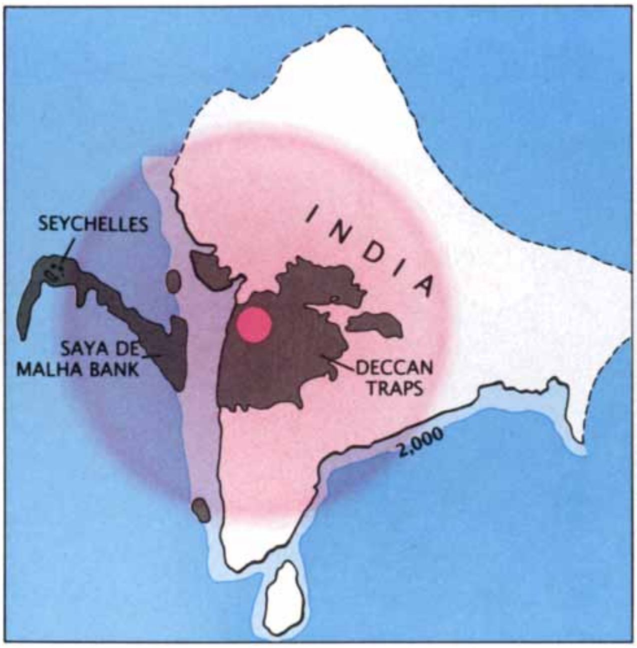
focused phenomena: narrow plumes of hot material triggering volcanism directly overhead. Yet the influence of the Cape Verde hot spot is apparent over a wide area. Maps of the seafloor and the earth's gravity field had revealed a 1,500-kilometer-wide swell in the seafloor surrounding the hot spot.

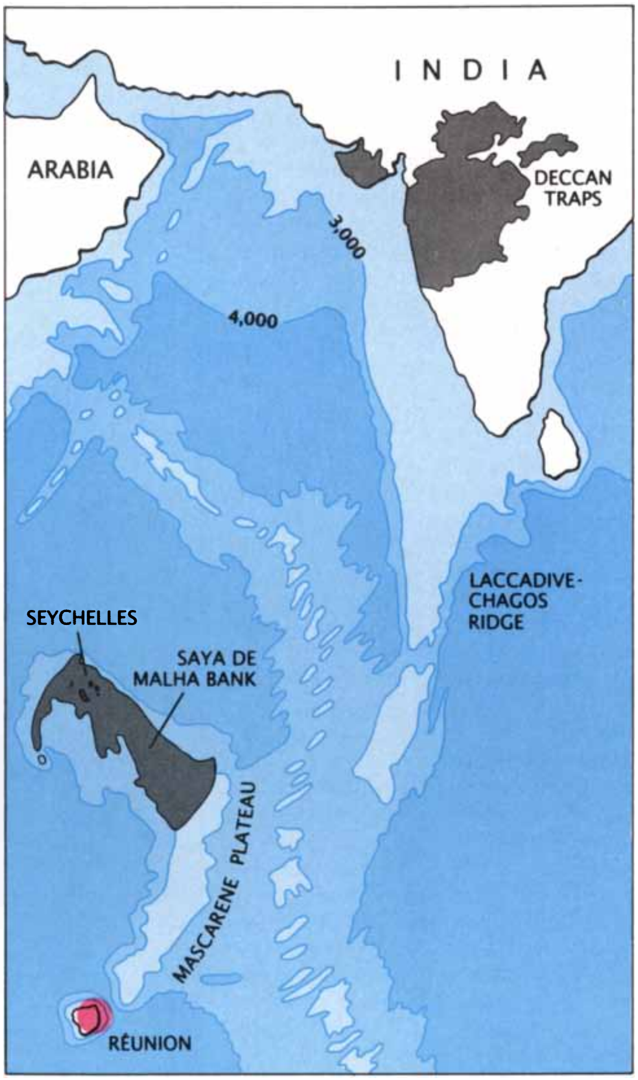
When we lowered instrumented probes to the seafloor to measure the heat flow through the sediments, we found the cause: well away from the hot spot the underlying mantle remains between 100 and 150 degrees hotter than normal. As the narrow rising plume under the Cape Verde hot spot meets the lithosphere, it is deflected into a mushroom of hot, buoyant material. The plume's broad cap is what uplifts the seafloor.

Similar regions of anomalously hot mantle are likely to surround other hot spots. The reason for massive volcanism at some rifts then becomes obvious. If a rift opens over a mushroom head of hot mantle, much more melt will be generated by decompression of the upwelling material than would result from rifting over normal-temperature mantle. The increase above the normal mantle temperature and the width of the anomalously hot regions are just what is needed to account for the volumes of melt released and for the length of the volcanic margins. The scenario also accounts for the speed of such eruptions. When unusually hot mantle wells up into a rift and undergoes a pressure reduction, melt will be generated immediately and will promptly bleed to the surface.

This explanation for volcanic rifts is easy to test: there should be evidence that a mantle plume was nearby each time a continental rift produced large amounts of melt. It is straightforward to reassemble rifted continents; the seafloor generated as they moved apart gives a continuous record of their movements with respect to the plate boundaries. Hot spots, for their part, are thought to be largely fixed in the mantle, but because plates themselves wander, a rift can migrate away from a hot spot. The relative motion of the hot spot is easy to trace because the plume leaves a trail of volcanic ridges, islands and sea mounts on the overlying plate.

The North Atlantic, for example, where submerged flood basalts were first mapped, has a plume under Iceland (a focus of intense volcanic activity). For 1,000 kilometers around Iceland the seafloor is about a kilometer shallower than normal, buoyed by the mushroom head of unusually hot mantle surrounding the hot spot. Reconstructions of the geography at the





INDIA AND THE SEYCHELLES had just rifted apart 65 million years ago (*left*). The volcanic rocks laid down during rifting include the Deccan flood basalts on land and underwater deposits that form the Saya de Malha Bank. (Extensive underwater deposits also probably lie off the Indian coast.) The volcanic margins were formed above the hot mantle surrounding the Réunion hot spot, which then lay near the rift. Since the rifting, plate motion has carried India northward; the volcanically thickened, shallow crust of the Laccadive-Chagos Ridge records the plate's passage over the hot spot (*right*). Some time ago the midocean ridge (the former continental rift) itself migrated over the hot spot; the gap between the Laccadive-Chagos Ridge and the Mascarene Plateau—the new hot-spot track—records the extent of sea-floor spreading since then.

© 1989 SCIENTIFIC AMERICAN, INC

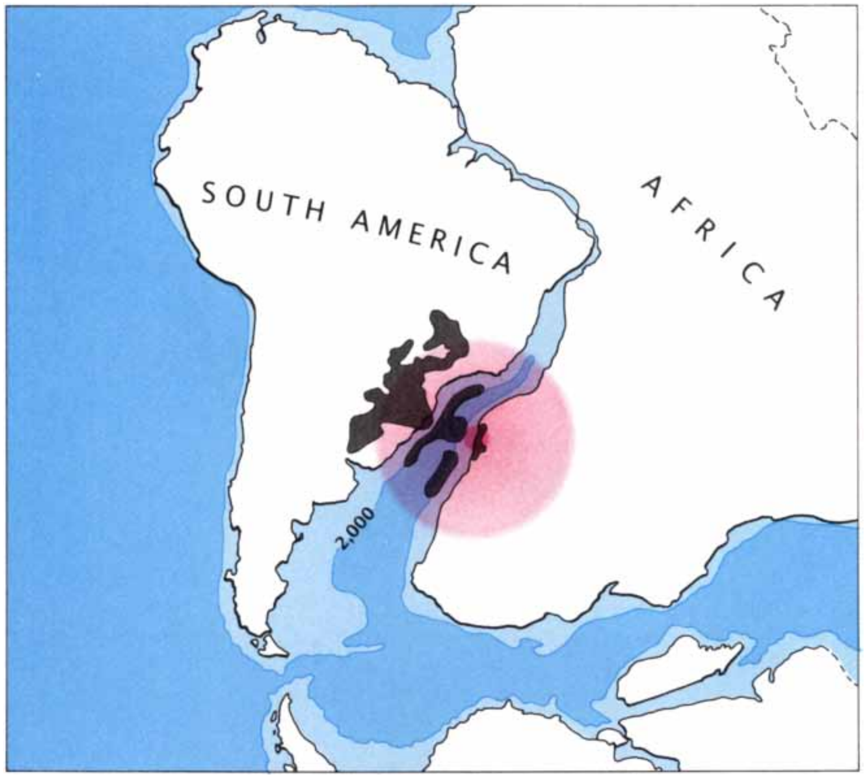
time the North Atlantic broke open, 57 million years ago, show that the plume then lay just to the west of the rift. Its head neatly encompassed the 2,000-kilometer extent of continental margins along which large thicknesses of volcanic rocks are now found.

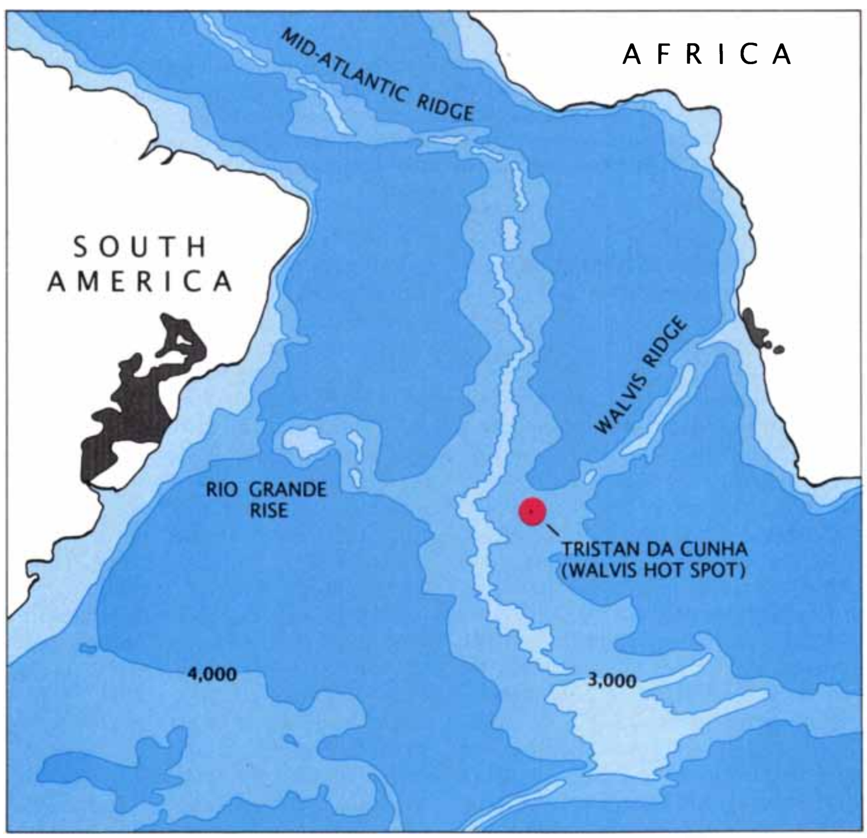
As soon as we were convinced that rifting above abnormally hot mantle could explain the extent of the volcanism in the North Atlantic, we realized that the scenario could also explain other features of the margins. Even though lithospheric stretching usually causes the crust to subside, for example, it was known from deposits of lake and river sediment trapped in igneous layers on the margins that the volcanism took place above sea level. What prevented the subsidence?

Part of the explanation is that the low-density plume uplifts the crust flanking the rift, as the swells around the Iceland and Cape Verde hot spots attest. Moreover, the added igneous material thickens the crust, which ordinarily would be thinned by stretching. Other factors contribute as well; all told, whereas crust stretched by a factor of five above normal-temperature mantle subsides immediately by two or three kilometers, the same amount of stretching above a 150-degree temperature anomaly actually leads to uplift. Of course thermal subsidence does eventually occur, which is why many volcanic margins now lie below sea level.

Rifting over high-temperature mantle also explains the composition of the volcanic layers. The laboratory results show that magma extracted at a higher temperature has more magnesium oxide than normal-temperature melt does. Accordingly, whereas magnesium oxide makes up about 10 percent of ordinary midocean-ridge basalts, some lavas extruded along a branch of the North Atlantic rift west of Greenland (but still within the compass of the Iceland plume) contain up to 20 percent magnesium oxide.

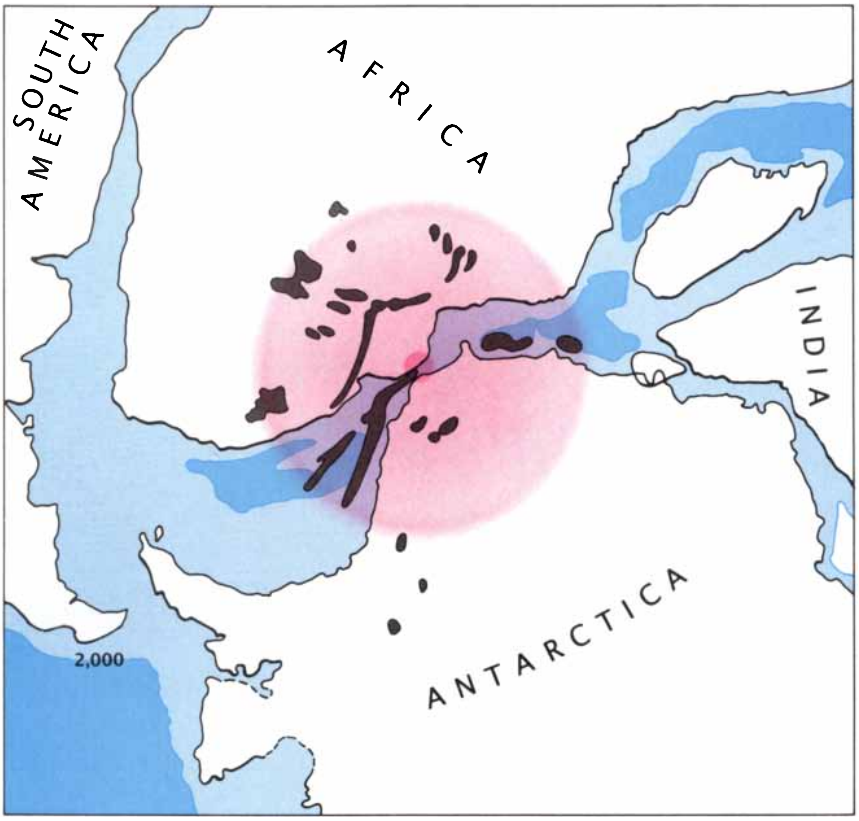
High-magnesium rocks in turn can explain certain seismic and gravity-field measurements along the Atlantic margins. Whereas the seismic velocity in the lower half of normal oceanic crust is 6.8 kilometers a second, the lower crust along the volcanic margins has significantly higher velocities in many places. Calculations predict that magnesium-rich igneous rocks intruded into the lower crust should have seismic velocities of more than 7.2 kilometers a second, about what is observed. Magnesium oxide also increases rock density, which can account for gravity measurements that





SOUTH ATLANTIC is shown 120 million years ago, just after the rift between South America and Africa had opened (*top*) and the ocean had begun to penetrate northward. The Walvis hot spot coincided with the rift; the associated region of hot mantle encompassed the volcanic margins that produced flood basalts during the continental rifting. The hot spot now lies under the island of Tristan da Cunha on the Mid-Atlantic Ridge (*bottom*); the meandering tracks the hot spot has left on both diverging plates in the form of thick volcanic ridges record the vagaries of plate motion.

© 1989 SCIENTIFIC AMERICAN, INC



BREAKUP OF GONDWANA, an ancient supercontinent incorporating many of today's landmasses, began 170 million years ago when a rift opened between Africa and Antarctica. Extensively eroded flood basalts on both continents and layers of basalt off the coast of Antarctica record the volcanism triggered by the breakup. A mantle plume and its associated thermal anomaly account for the volcanism's extent.

require unusually high crustal densities along such margins.

Continental Flood Basalts

Our scenario for the origin of volcanic margins also fits the evidence outside the North Atlantic. According to reconstructions of landmasses and hot-spot positions at the time of rifting, other rift margins that are edged by massive volcanic deposits also formed above mantle thermal anomalies. In accounting for the volcanic margins, the model simultaneously explains the extensive flood basalts that often stretch inland from them.

Sixty-six million years ago, for example, the largely submerged continental block that bears the Seychelles islands rifted away from western India. Like the young North Atlantic, this rift extruded vast amounts of magma onto the continental margins; the volcanic rock now forms a submarine plateau known as the Saya de Malha Bank, east of the Seychelles, and extensive deposits along the west coast of India. The melt also poured hundreds of kilometers inland, forming a vast basalt plateau called the Deccan Traps.

After the rifting the Indian subcontinent continued to drift northward on a collision course with Asia. It left the plume responsible for the volcanism far to the south, where it now lies under the volcanic island of Réunion. A trail of volcanic islands, sea mounts and ridges left on the oceanic crust that has moved across the hot spot shows that 66 million years ago it lay directly under the developing rift.

Comparable flood basalts extend inland from parts of the Atlantic coasts of South America and Africa; along these coasts, seismic profiles reveal the igneous layers typical of a volcanic margin. The South Atlantic began to open between the two landmasses at their southern end about 125 million years ago. Like a zipper being opened the split moved steadily northward and was completed in about five million years. Throughout that period a thermal anomaly called the Walvis hot spot lay directly under the central part of the rift. Decompression over the thermal anomaly released massive quantities of lava onto Africa and even more—perhaps two million cubic kilometers—onto South America, where it filled valleys as deep as 300 meters before overtopping them and flowing onward. The timing of the eruptions is not as well known here as it is for the Deccan, but the volcanism seems to have lasted only a short time and to have reached its peak from 121 to 120 million years ago, just as the landmasses separated.

A still older example of flood basalts associated with volcanic rifting dates from the breakup of Gondwana, the supercontinent that included Africa, Antarctica, South America, India and Australia. About 170 million years ago a rift appeared between Africa and Antarctica, directly over a plume that now lies in the southern Indian Ocean. Decompression melting created the 10-kilometer-thick volcanic sequences that have recently been discovered on the margin of Antarctica, and the existing flood basalts show that the magma poured far inland as well.

Plumes and Rifting

We have assumed that these vast quantities of magma are generated passively, in response to the stretching and rifting of the lithosphere. Actually the mantle does play an active role: for so much melt to be produced a plume must have brought up unusually hot material from deep in the mantle. Could plumes, in addition, somehow drive the rifting process?

J. Tuzo Wilson of the University of Toronto was the first to suggest, more than 20 years ago, that hot spots might somehow be responsible for splitting the lithosphere and initiating rifts. Yet melt actively forced upward from a plume into the lithosphere cannot be what starts a rift, since extensive melting does not begin until the mantle has welled up above the base of unrifted lithosphere. Moreover, hot spots clearly are not crucial for rifting: normal, nonvolcanic margins such as the Biscay margin were formed well away from any mantle plume.

Even if plumes do not drive rifting, the one or two kilometers of uplift they cause might help it along by way of gravity, which enables the rifting plates to slide apart down opposite sides of the swell. The advent of a plume may thus have been the deciding factor in starting rifts. In the North Atlantic, for example, stretching and small-scale rifting took place for many tens of millions of years before a fully oceanic rift developed between Greenland and Northern Europe about 57

© 1989 SCIENTIFIC AMERICAN, INC

million years ago. The arrival of the Iceland plume from deep in the mantle may have been the trigger. The earliest signs of the plume—lava flows in northwestern Great Britain—predate the breakup by only five million years. (The plume's advent is dramatically recorded on the other side of Greenland, in the Davis Strait and Baffin Bay. A rift had been quietly opening there for 30 million years, but about 58 million years ago it abruptly produced thick igneous sequences as the mushroom of hot material arrived.)

If a new plume sometimes precipitates rifting, the aging of the plume could explain the disparity in thickness between most rifted margins and the adjacent oceanic crust. In the North Atlantic, for example, the igneous layers on the rifted margins are more than 15 kilometers thick. They grade into oceanic crust, generated after the rift opened, that is between 10 and 12 kilometers thick—thicker than normal oceanic crust but thinner than one might expect if the mantle temperature had remained constant. The Iceland hot spot, after all, still sits directly on the midocean ridge.

The falloff in crustal thickness, or, equivalently, in the amount of magma released at the rift, probably reflects the structure of the plume. Computer modeling suggests that mantle plumes start with a blob of material perhaps 50 or 100 degrees hotter than the later, steady-state flow. In addition, the blob, which forms the head of the mushroom, has a larger radius than the plume that follows. The higher temperature and larger volume of the material reaching the top of the mantle as the rift opens should enhance early melting and account for the exceptional thickness of the igneous margins. Thinner oceanic crust should follow as the initial hot blob gives way to a steady plume flow.

The Growth of Continents

So much melt is released each time a continent breaks up near a hot plume that volcanism at rifts must play a much larger role in the evolution of the earth than had been thought. The process may, for example, offer a partial answer to the question of how the continents were built. Our model of melt production, coupled with the evidence from the North and South Atlantic and India, shows that some 10 million cubic kilometers of molten rock are released every time a continent rifts apart near a hot plume. By flowing onto the continental crust and underplating it, the igneous material increases the volume of the continental fragments.

We estimate that an episode of volcanic rifting has taken place once every 30 million years, on the average. Hence, the process may have been adding an average of one third of a cubic kilometer of material to the continents each year in recent eras. Early in the earth's history the mantle must have been hotter than it is now, and rifting and decompression would therefore have generated even more melt. The average rate of continental growth since the earth was formed has been perhaps one cubic kilometer a year, and so volcanic rifting may have been the chief source of early continental growth. The main problem with the proposal is the high magnesium oxide content of melts released in volcanic rifting: it is not consistent with the observed composition or seismic properties of continental crust.

Whatever their precise role in the history of the earth, these volcanic outbursts can now take their place along with other geologic processes in the framework of uniformitarianism: the notion that the geologic past can be explained in terms of the same phenomena now shaping the earth. The unmatched scale of certain eruptions, notably the Deccan basalts, has led some workers to invoke causes outside normal earth processes—impacting asteroids, for example. But we do not think such catastrophes are required. Thick marine sequences of igneous rock, flood basalts on land and perhaps even mass extinctions—all can be explained by the interaction of familiar, ongoing earth processes.

FURTHER READING

ANOMALOUS HEAT FLOW AND GEOID ACROSS THE CAPE VERDE RISE: EVIDENCE FOR DYNAMIC SUPPORT FROM A THERMAL PLUME IN THE MANTLE. R. C. Courtney and R. S. White in *Geophysical Journal of the Royal Astronomical Society*, Vol. 87, No. 3, pages 815–867; December, 1986.

MAGMATISM AT RIFTED CONTINENTAL MARGINS. Robert S. White, George D. Spence, Susan R. Fowler, Dan P. McKenzie, Graham K. Westbrook and Adrian N. Bowen in *Nature*, Vol. 330, No. 6147, pages 439–444; December, 1987.

THE VOLUME AND COMPOSITION OF MELT GENERATED BY EXTENSION OF THE LITHOSPHERE. D. McKenzie and M. J. Bickle in *Journal of Petrology*, Vol. 29, No. 3, pages 625–679; June, 1988.

MAGMATISM AT RIFT ZONES: THE GENERATION OF VOLCANIC CONTINENTAL MARGINS AND FLOOD BASALTS. R. White and D. McKenzie in *Journal of Geophysical Research*, Vol. 94, No. B6; June, 1989.

WORLD'S SMALLEST——
——WEATHER STATION

THE AMAZING WEATHER COMPUTER THAT YOU CAN HOLD IN THE PALM OF YOUR HAND.

DIGITAR's new TWR-3 Micro Weather Station includes a computer, precision wind vane and speed sensor with mounting hardware, and 40 feet of cable. For only $159.95. With the optional, automatic-emptying RG-3 Rain Collector ($49.95) you can even monitor rainfall!

- WIND SPEED
- WIND DIRECTION
- WIND CHILL
- WIND GUST RECORD
- TEMPERATURE
- HI/LOW TEMP RECORD
- RAINFALL (Optional)
- TIME OF DAY
- AUTO SCAN
- METRIC/STANDARD
- NICAD READY
- ONE YEAR WARRANTY

Add $5.80 for shipping, plus sales tax for California deliveries.

DIGITAR

MADE IN U.S.A. 3465 Diablo Avenue Hayward, CA 94545 M/C & VISA

ORDERS ONLY: 1-800-678-3669
(7 a.m. - 5:30 p.m. Pacific Time)

INFORMATION: 415-732-7814 FAX: 415-732-9188

Want to brush up on a foreign language?

With Audio-Forum's intermediate and advanced materials, it's easy to maintain and sharpen your foreign language skills.

Besides intermediate and advanced audio-cassette courses—most developed for the U.S. State Dept.—we offer foreign-language mystery dramas, dialogs recorded in Paris, games, music, and many other helpful materials. And if you want to learn a *new* language, we have beginning courses for adults and for children.

We offer introductory and advanced materials in most of the world's languages: French, German, Spanish, Italian, Japanese, Mandarin, Greek, Russian, Portuguese, Korean, Norwegian, Swedish, and many others.

Our 17th year

CALL 1-800-243-1234 FOR FREE 32-PAGE CATALOG, OR WRITE:

AUDIO-FORUM®
THE LANGUAGE SOURCE Room C724,
96 Broad St., Guilford, CT 06437
(203) 453-9794

© 1989 SCIENTIFIC AMERICAN, INC

SCIENCE AND BUSINESS

High-tech texts, Taguchi methods, a nose for wine, inflated measures

Show and Tell

Hypermedia turns information into a multisensory event

A high school student, assigned a report on women in American history, turns to a computer system tied to a video monitor and watches a short film on immigration. Using a computer mouse, she stops the video on a picture of a woman; a "notecard" with information on women immigrants appears on her computer screen. Her curiosity piqued by a reference to the garment industry, the student locates another clip on industrialization and sees pictures of women in sweatshops. Along the way she collects snippets of the images in an electronic file to supplement the report she will write. Elsewhere, a telephone company executive weighs what new services to offer to clients. His computer illustrates each of the services under development at Bell Communications Research (Bellcore); when prompted, it offers more detail—notecards about the software and technology needed to implement a service or the likely costs and advantages.

These snapshots illustrate hypermedia, a new way to manage information. The American history project will be released by *National Geographic* (in conjunction with Lucasfilm) in January; several telephone companies already use Bellcore's package. Just more data bases? No, say software developers. Hypermedia is a different way of organizing information in many media—text, pictures, video clips and sound. It relies on computers, videodisks and monitors to create a multimedia library, tailored to a user's needs.

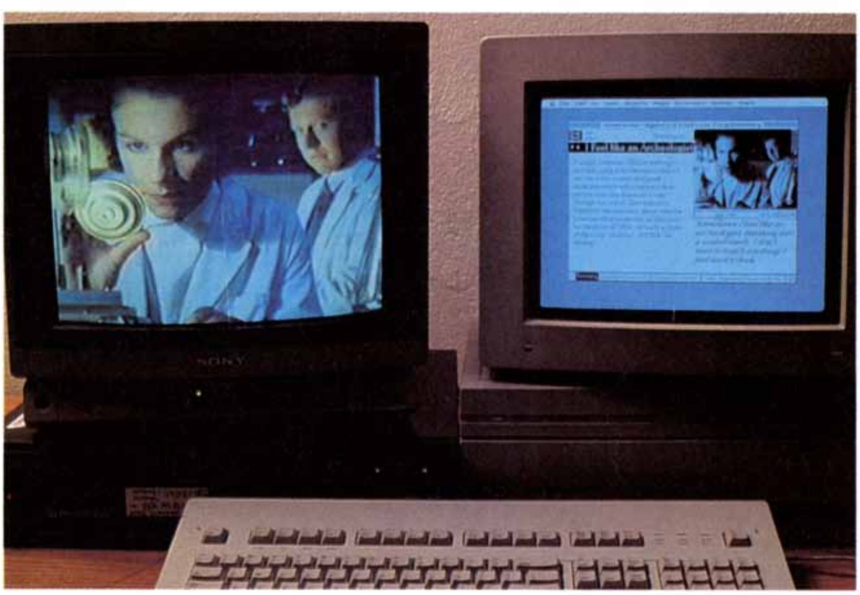
Hypermedia systems range from classroom products drawing on fixed, limited data bases to complex programs for managing changing streams of information in business and technology. Their common thread is nonlinearity—in effect, the lack of a single thread. Rather than leading a person straight through the information, hypermedia allows each user to cut his or her own path. More fundamentally, hypermedia is another step away from systems for the computer intelligentsia and toward systems that better accommodate how most people think.

The inspiration for hypermedia was a 1945 article in the *Atlantic Monthly* written by Vannevar Bush, who directed the U.S. Office of Scientific Research and Development during World War II. Bush predicted that science would develop a new approach to organizing information. In his scenario, a "memex"—a sort of intelligent agent—would help people to create customized data bases.

The technical foundations for hypermedia were laid in the late 1960's. Yet it took relatively cheap, powerful computers and data-storage technology (such as videodisks) to make hypermedia practical. In 1987 Apple Computer released its "HyperCard" software, which quickly became a platform for a host of projects, including those of *National Geographic* and Bellcore. HyperCard works much like an electronic index-card file. Facts are stored on electronic notecards that can be shuffled into any desired order.

Tying cards to slides, video clips and sound—the essence of these early hypermedia projects—can transform classroom lessons, says Kristina Hooper, director of the multimedia laboratory at Apple; students can, for example, read notes about a scientific discovery, hear an interview with the lead scientist and explore diagrams that detail the finding. One of the first products is a package on the 1988 election created by ABC *News* and Optical Data in Warren, N.J.; another, on the conflict in the Middle East, is slated to be available this summer.

John Steinbeck and Shakespeare are also being adapted to hypermedia. Along with reading *The Grapes of Wrath,* students can learn about the Great Depression by tapping into fireside chats by Franklin D. Roosevelt, crop statistics and Woody Guthrie songs. Students of Shakespeare can "stage" his plays electronically, drawing on a wide cast of characters stored on videodisk. For medical students, the "Electric Cadaver" at the Stanford University Medical Center offers a tour of anatomy. Start with a full skeleton, then examine a shoulder joint. Or



HYPERMEDIA can link video to a data base of facts. In this presentation on DNA, developed jointly by Apple Computer, Lucasfilm and the Smithsonian, students can learn the science behind the story by stopping the video and exploring the computer data base. Photograph by Ed Kashi.

© 1989 SCIENTIFIC AMERICAN, INC

SCIENCE/SCOPE®

A single, wide-band voice, data, and image traffic system will improve communications and reduce project costs at Hughes Aircraft Company. The Integrated Digital Network (IDN) brings digital communications circuits to all major company sites. These circuits are more reliable than analog circuits, and new technology can be more readily incorporated into a digital network. IDN will allow users to reduce communications costs by combining voice and data traffic over the same lines. The system's extremely wide bandwidth will also permit the introduction of new services, such as video teleconferencing, which can result in reduced travel costs and increased productivity.

A computerized, voice-output system will automatically pick the proper radio network and frequency, and talk back to the air crew of an advanced helicopter. The Communications and Identification Subsystem, under development at Hughes for the U.S. Army's Light Helicopter Experimental (LHX) Program, also includes an over-the-horizon, high frequency communications radio with an anti-jam feature. The subsystem has a number of radio frequencies that can be used under a variety of conditions, including ultra high frequency and very high frequency FM and AM channels. By switching channels, the subsystem reduces the chances of enemy jamming interrupting critical communications.

A diamond film deposition system that allows an exceptionally high degree of control over the film deposition process has been demonstrated by Hughes. The system successfully deposited small polycrystalline diamond islands on a silicon substrate at temperatures below 200 degrees Centigrade. This overcomes a principal impediment to diamond-film applications, namely the lack of a way of depositing quality films at reasonably low temperatures. Thin films of diamond can serve a wide variety of applications because of their hardness, infrared transparency, high thermal conductivity, and high temperature operation. In addition, their excellent semiconductor properties promise performance superior to gallium arsenide.

Artificial intelligence (AI) techniques have reduced fault-location time by 94 percent in the testing of complex infrared sensors during assembly. Working with a detector design, manufacturing expert, and actual test data, Hughes has developed a set of rules and procedures for an expert system called Coldfinger. It correctly detected and located six types of problems in an infrared sensor containing 160 detector elements and complex processing electronics. One set of defects took a human expert 30 minutes to define, but was identified by Coldfinger in just two minutes. Hughes is also developing AI systems for autonomous land and underwater vehicles, and diagnostic techniques for other manufacturing operations.

Hughes is seeking experienced engineers and scientists to design, develop, and produce Hughes' new line of body-stabilized HS 601 communications satellites. Openings are in the fields of: software, computers, and data processing systems; electrical components; microwave/RF communication systems development; on-board spacecraft electronics and control systems; satellite design, integration, propulsion, and electrical power system development; spacecraft manufacturing, systems test and evaluation; GaAs applications R&D. Send your resume to Michael Martinez, Hughes Space & Communications Group, Dept. S2, S4/A300, P.O. Box 92919, Los Angeles, CA 90009. Equal opportunity employer. U.S. citizenship may be required.

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

HUGHES

Subsidiary of GM Hughes Electronics

©1989 Hughes Aircraft Company

© 1989 SCIENTIFIC AMERICAN, INC

stimulate a nerve and watch the effect on other body parts. "When you reach a critical mass of interrelated material, this kind of system is extremely useful," says Steven J. Freedman, a developer of the project. (Freedman expects the "Electric Cadaver" to be on the market in 1990.)

Critics argue that splashy videos and other media can add more gimmickry than substance. Computer-generated multimedia presentations may merely "imitate television's habit of creating increasingly stimulating experiences rather than imparting knowledge," writes Nick Arnett, editor of *Multi-Media Computing & Presentations.* But educators have become increasingly sensitive to students who do not readily learn from print. If they can absorb information more easily from hypermedia, the new technology will be a valuable classroom tool, observes Mary Alice White of Teachers College, Columbia University.

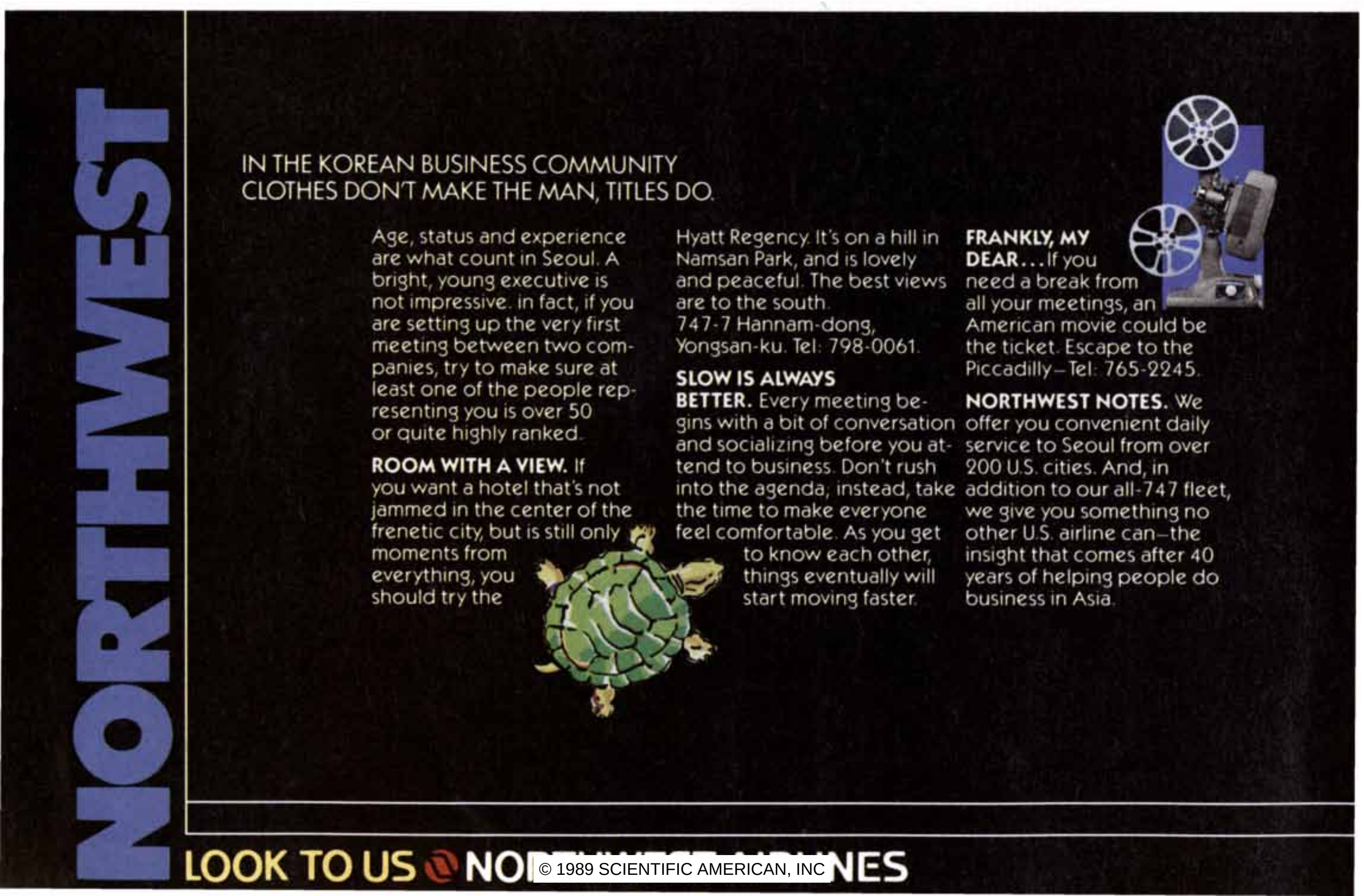
Other researchers have moved beyond cards and are trying to build hypermedia systems that make full texts—such as manuals and books—more accessible. People searching for information, even thumbing through a book, immediately locate what they need only about 20 percent of the time, observes Thomas K. Landauer, a Bellcore manager. So Bellcore's "SuperBook" features particularly flexible searching techniques. (Given a key phrase, it will identify sections even when the words in the phrase are out of order.) SuperBook also condenses chapter headings before and after a selected section to give a "fish-eye" view of the surrounding material.

Organizing chunks of information into nonlinear webs also appears to be well suited for managing highly complex projects, researchers say. The Microelectronics and Computer Technology Corporation (MCC), for instance, is using hypermedia in a system for managing the production of new software. Several years ago officers on board the USS *Carl Vinson* used an experimental hypermedia system built at Carnegie Mellon University to help manage the aircraft carrier; Westinghouse has similarly explored how hypermedia might provide nuclear-power plant operators with easy access to information during emergencies. Knowledge Systems in Pittsburgh now sells a commercial package based on those prototypes.

Many of these hypermedia systems incorporate other software tools, including expert systems. Building on Knowledge Systems's software, Bellcore is creating a hypermedia package to help regional companies develop business plans and oversee their subsidiaries. If a company adopts rules specifying that its subsidiaries should not compete with one another, the system can flag the conflict for managers. The benefit of hypermedia? Conventional expert systems are best used by those trained in computer science; hypermedia can be understood by anyone, say Bellcore workers.

Yet hypermedia designers still face many issues. Is there enough information—both text and video—available for comprehensive presentations? (Even *National Geographic* and ABC have had to seek additional sources.) Will laser disks become obsolete in the face of newer storage technologies? What features will make hypermedia systems valuable and easy to use?

What is clear is that hypermedia is not a single invention—like the transistor—that can be easily copied and exploited by many companies. "Each company will have to work out how to use hypermedia in a way specific to its own need to organize information," says David L. Turock, a Bellcore manager. Those companies that grapple early with how to structure information, he predicts, will build the strongest leads in the information race. —*Elizabeth Corcoran*



© 1989 SCIENTIFIC AMERICAN, INC

Quality Conscious
Even techniques for quality are imported

Genichi Taguchi is probably amused at the irony. More than 30 years ago U.S. statisticians lectured Japanese companies about quality manufacturing. Today Taguchi's quality-design philosophies draw followings in corporate America.

Quality control is far from new in the U.S.; analytical methods for monitoring quality were developed in the 1920's. Among the early legends was statistician W. Edwards Deming, who introduced quality monitoring during production to Japanese companies. Japan's "Deming prize" for quality soon became one of the most prestigious honors. In the U.S., it took the rude shock of losing sales to foreign competitors in the early 1980's to convince many manufacturers to embrace Deming and to adopt Japanese-like attention to quality.

Now U.S. companies are rushing to implement stronger quality-control measures from a product's initial design through production. (Inspecting for defects at the end of the production line is still an important, albeit costly, quality measure.) Over the past two years managers and workers at McDonnell Douglas Missile Systems have received pocket-size cards on quality; among the guides included were Deming's "14 points for transforming industry" and Taguchi's principles for quality design.

What is the Taguchi method? Taguchi, a statistician and industrial consultant, proposes designing "robust" products—ones that can remain stable through long and hard use by the consumer and in spite of variability in the manufacturing process. Suppose a company makes tiles, explains Duane E. Orlowski, technical editor of the *ASI Journal,* which is devoted to furthering Taguchi's methods. And suppose further the tiles are irregular because the kiln does not heat uniformly. "You can't control these kinds of 'noise' factors," Orlowski says. Instead Taguchi recommends reconsidering the design of the product—changing the specifications or the materials—to ensure that subsequent tiles will be uniform regardless of the noise.

Taguchi's robust products depend on "experimental designs," a concept pioneered in England in the 1920's, notes James P. Pennell, a researcher at the Institute for Defense Analyses. Researchers pick a quality trait they wish to measure (perhaps the precision of tile dimensions) and select those product factors that affect the trait most directly. Multiple factors are systematically and simultaneously altered, and the resulting products are measured. By assembling the data in orthogonal arrays, workers can infer the effects of many combinations of factors. In this way the number of experiments needed for meaningful results is reduced. "We will deliberately vary as many as 10 or 20 factors in a preplanned way so that the separate effects and interaction of variables can be measured and estimated," says Donald W. Marquardt, manager of the applied statistics group at DuPont.

Experiments can pinpoint features less critical to the product's performance, thereby allowing producers to substitute less expensive materials or techniques and trim unit cost. "Taguchi argues that lower cost is the driving force behind higher quality," writes Lance A. Ealey in his book, *Quality by Design,* which focuses on Taguchi. "The real purpose of technology... should be to let unit manufacturing cost take precedence over quality," Taguchi says. "If Japanese automobiles cost two times more, most Americans would stop buying them."

In the U.S., the Taguchi principles have been credited with a number of

© 1989 SCIENTIFIC AMERICAN, INC

successes. A division of AT&T, for example, used Taguchi methods in a photolithography application and reduced variance in the process fourfold, "fatal defects" threefold and processing time twofold. Aerojet Ordnance used Taguchi tests to create a data base on ammunition and so was able to verify the performance of new products in about a month, a 98 percent improvement over its previous practices, according to Aerojet.

Taguchi is not without his critics. Statisticians argue heatedly over his methodologies, in particular the way he calculates robustness based on a "signal-to-noise" ratio. "There's no question Taguchi has made significant contributions," Marquardt says, "but we look on [his] specific technical approaches as ones we may or may not include." According to Pennell, a tour of Japanese companies that had won the Deming prize revealed that they were familiar with Taguchi methods and used experimental designs but that few applied his techniques.

Applying every analytical quality tool in the book is not enough, stress the experts. "Everybody knows what needs to be done," Pennell says. "But if the program isn't supported by management, it falls into disuse." —*E.C.*

A Connoisseur's SNIF

Magnetic resonance aids in identification of wines

Winegrowers often quarrel over fine points of grapes, weather and vinting techniques. Now a chemist has found a way to inject data into the squabbles.

Gerard Martin, a chemist at the University of Nantes, has developed a process called site-specific natural isotope fractionation (SNIF), which uses magnetic resonance imaging to analyze the ratios of different hydrogen isotopes in simple molecules. SNIF can detect additives to wine, such as sugar used to boost fermentation (a process called chaptalization), and help track down the homelands of mislabeled wines. Martin's son, Gilles, has consequently formed a company called Eurofins, which is also putting SNIF to work analyzing other alcohols, fruit juices and flavorings.

The European Community has, in turn, greeted SNIF enthusiastically. According to officials, fraudulent labeling—wines with labels that boast a single, prestigious origin when in fact their grapes came from several countries—costs consumers several hundred million dollars annually. Moreover, chaptalization, which is particularly useful when grapes do not ripen enough to ferment properly, vexes European winegrowers favored by sunny climates. Whereas Italians and Greeks refer to chaptalization as "adulteration," their French and German competitors call it "enrichment," points out wine aficionado Robert W. Cahn in *Nature.* (In the U.S., chaptalization is prohibited only in California.)

In recent years analysts have turned to mass spectrometry, which can detect wines boosted with cane sugar based on the sugar's distinctive carbon isotopes. But chaptalization is also done with beet sugar or concentrated grape sugar. It is the beet sugar that irks the purists. Because beet sugar and wine have identical carbon isotopes, there has been no reliable way to detect many chaptalized wines, explains Richard Gahagan, wine technical advisor to the U.S. Bureau of Alcohol, Tobacco and Firearms.

SNIF relies not on carbon isotopes but on the ratio of ordinary hydrogen to its heavy isotope, deuterium. Beet sugar has its own distinctive deuterium-hydrogen ratio that SNIF can pick out directly. Deuterium-hydrogen ratios in grapes also vary significantly according to latitude, weather conditions and type of grape, Martin reports. For instance, grapes grown in southern Italy can have a 5 percent greater concentration of deuterium than those from Germany, he says; hot, dry conditions also contribute to higher deuterium concentrations.

To trace the origins of a wine suspected to be a blend, Eurofins applies SNIF as well as more traditional tools, including mass spectrometry. The company has already compiled a data base of the characteristics of some 4,000 samples of wine from different growing areas. Martin says that SNIF can easily discriminate between a Bordeaux and a Burgundy; distinguishing one variety of Bordeaux from another is more difficult, he concedes. Although the analysis is not cheap, ranging from $200 to $300 per sample, Martin predicts the cost will drop to $50 or so within four years.

The Commission of the European Communities has officially recognized SNIF as the way to monitor chaptalization. Moreover, SNIF makes it technically feasible for the commission to harmonize the differing national regulations on chaptalization—perhaps, Gahagan suggests, by sanctioning only the use of concentrated grape sugar. In the event of a single European policy on chaptalization, he adds, the U.S. government might also turn to SNIF to ensure that wine shipped to and from Europe conformed to the commission's regulations.

Yet many political and cultural obstacles to a single chaptalization policy remain. Wine connoisseurs, for their part, seem unruffled by the debates and, more significantly, by chaptalization. "I can taste it sometimes, if people point it out to me," says David Rowe, editor of *Decanter.* James Laube, an editor for the *Wine Spectator* concurs: "When grapes are closer to maturity, it [chaptalization] might even enhance the flavor." —*E.C.*

THE ANALYTICAL ECONOMIST

The CPI should measure inflation. Your mileage may vary.

Almost everyone worries about inflation. When it is out of control, money printed in the morning may be worthless by noon; on the other hand, stable or declining prices may signal a stagnant economy. The U.S. Federal Reserve raises interest rates to check inflation and lowers rates when the monster appears quiescent.

The standard indicator of inflation is the consumer price index—the average cost of a particular "market basket" of goods and services. Every month bond traders and stockbrokers wait anxiously to see whether the newly announced CPI figure matches their expectations; once a year Social Security administrators, IRS accountants and personnel departments nationwide refigure entitlements, tax brackets and wages based on CPI changes. More than 60 million people receiving government benefits are directly affected by the published figure. "It's assumed the numbers are reliable," says Craig Coats, vice-chairman of the securities firm Vouté Coats Stuart & O'Grady. Traders react "quickly and strongly" to changes in the CPI as well as to other economic statistics.

Such a powerful number might be expected to enjoy as much objective reality as a thunderstorm. In fact, the CPI does not necessarily correspond at all to inflation as many people experience it. To calculate the CPI, fieldworkers from the Bureau of Labor Statistics visit representative stores throughout the U.S. every month and get price quotes for items ranging

© 1989 SCIENTIFIC AMERICAN, INC

from single cans of soda to pickup trucks. They adjust the quotes to account for changes in quality or quantity and then weight them according to the proportion of household expenditures each item represents (as determined by surveys and adjusted about every 10 years). The resulting sum is the consumer price index. For the mythical "average" household, CPI increases will match inflation. Almost everyone else will either gain or lose a little—or perhaps a lot.

The CPI has flaws both in execution and in concept. Like any survey, it has biases and sampling errors: the results will never exactly match the "true" average of prices nationwide. The BLS used to estimate that the monthly overall CPI was within .1 percent of the true figure, but it stopped estimating errors in 1978 because its research indicated the real errors were probably at least double the estimates. The BLS says it hopes to begin issuing revised error figures in 1990.

The old methods did not account properly for the interactions between sampling errors for different items, according to chief price statistician Curtis Jacobs; nor did they properly assess the effects of sampling errors for individual CPI categories on the overall error. As a result, a 6.0 percent annual rate of increase for the CPI could potentially correspond to a true rate anywhere between about 3.5 and 8.5 percent. Thus, the monthly CPI may not accurately distinguish between a 4 percent inflation rate and a 7 percent rate—a difference that could send securities markets soaring or plunging.

Measuring price increases for longer periods—three months or a year, as is done for cost-of-living adjustments—reduces the uncertainty: a report that the CPI is up 5.0 percent from a year ago very probably places the true figure between 4.6 and 5.4 percent.

Jacobs and his colleagues also are working to determine whether initial biases in measuring spending patterns may contribute even more to CPI errors than do monthly variances in price quotes. The information the consumer-spending surveys provide is uncertain, and it grows increasingly outdated in the period between surveys. Changes in consumers' spending habits may alter actual cost increases substantially. For example, although the CPI measured oil and gas prices rising at more than 12 percent a year from the early 1970's to the early 1980's, consumers spent only about 9 percent more a year on oil and gas. By 1983 the CPI was overestimating expenditures by about 40 percent.

Moreover, during the 1970's CPI increases varied by as much as 10 percent among demographic groups, because the groups spent money on different mixes of goods and services, according to BLS economist Mary F. Kokoski. Lower-middle-class families, for example, saw prices increase an average of 8.3 percent a year while upper-middle-class singles saw annual increases of 7.5 percent.

Such variations are particularly important for those whose income—say, from Social Security checks—is indexed to published rates of inflation. Senior citizens spend more of their money on medical costs than do most consumers, and so they might be more sensitive to the roughly 7 percent rise in medical prices last year than to the roughly 4 percent CPI increase.

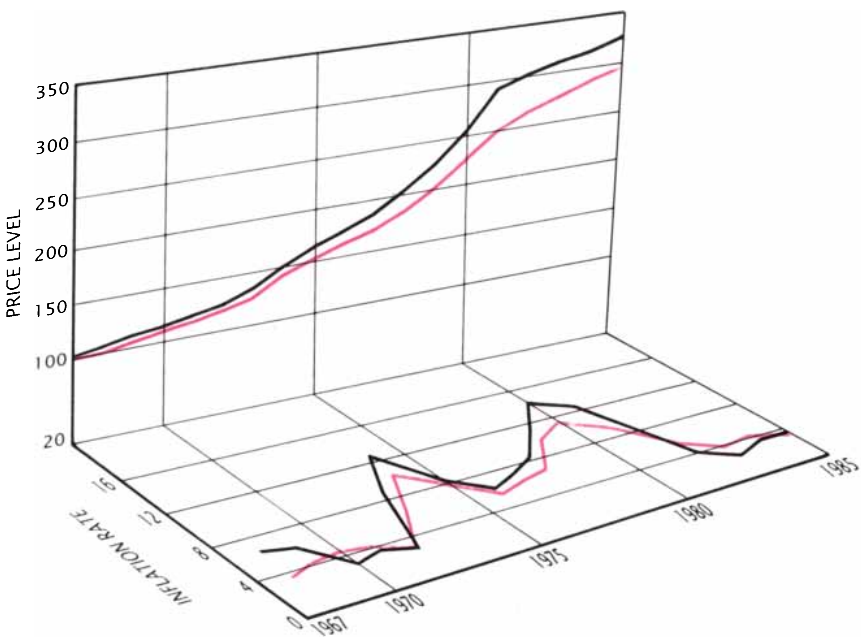
Even more important for the elderly, the BLS gathers data only on "noninstitutional" costs of living, so people living in nursing homes are essentially ignored when it comes to adjusting benefits or pensions. Although the BLS has looked into specially tailored price indexes, Kokoski says, differences in the spending patterns within a demographic group are generally at least as large as differences between groups, so tailored indexes could be as inaccurate as the general index.

And even if it were possible to formulate a consumer price index (or a family of indexes) that accurately reflected spending patterns, that would still not answer a question many people ask: "How much income do I need to maintain my standard of living?"

For that, Kokoski and her colleagues have constructed an index that measures changes in how much a household must earn before taxes to pay for the standard consumer price index market basket. During periods of high inflation (such as the late 1970's), the TPI (tax and price index) outpaces the CPI; between 1975 and 1980 even those whose wages were indexed to inflation lost an average of six cents on every dollar of earnings. In the early 1980's, thanks to reductions in tax rates, the TPI fell below the CPI, and taxpayers with indexed incomes actually gained a fraction of a cent.

Beyond adjustments for taxes and spending patterns lies a more fundamental question: what constitutes a standard of living? Should, for example, increased costs for entertainment be considered roughly as important as those for medical care, as the weighting system now makes them? Or should standard-of-living measurements take into account intangibles such as crime rates or pollution? Many economists have tried over the years to come up with such quality-of-life indexes, but no index has been widely accepted. Only one thing seems certain: if you have a nagging feeling your income is not keeping up with inflation, you are probably right.

—Paul Wallich and Elizabeth Corcoran



CONSUMER PRICE INDEX* (red) *measures only the cost of goods. (The top graph shows costs; the bottom graph shows the annual increases as a percentage.) To gauge the pretax earnings needed to purchase those goods, the Bureau of Labor Statistics compiled an experimental Tax and Price Index* (black). *Indexing of tax brackets may have brought the CPI and TPI into step.

© 1989 SCIENTIFIC AMERICAN, INC

The Gastrointestinal Tract in Growth and Reproduction

It is the largest endocrine gland in the body, and it has a significant role in the readjustment of metabolism that accompanies pregnancy as well as fetal and infant growth

by Kerstin Uvnäs-Moberg

The gastrointestinal tract may seem to be simply a long, winding cavity into which food is introduced, there to be digested into nutrient molecules that can be absorbed through its walls and enter the circulation. In reality the gut is much more than that. It is the largest endocrine gland in the body, and the hormones it secretes exert profound effects not only on the process of digestion itself but also on the metabolism of ingested nutrients and even on the emotions and behavior.

At no stage of life are these physiological functions more critical than during growth and reproduction. Any organism needs more nutriment when it is growing than when it is not. The young of many species eat (relatively) more than adults do; the calorie intake of human infants per kilogram of body weight is more than four times that of adults. The demand for food is also high in organisms undergoing reproduction, which is often preceded by a period of increased uptake and storage of energy. So important is nutrition to reproduction that reproduction simply does not take place in the absence of adequate food.

In mammals, including human beings, most reproductive work is done by the female. Women therefore differ from men in that rather than having only one major period of growth, from infancy through adolescence, they may grow again one or more times as adults: during pregnancy. Beginning early in pregnancy a woman gains weight, storing energy as fat against the demands of the fetus and in preparation for the heavy demands that will come with lactation and breastfeeding. And in human beings, too, energy intake and storage are related to the ability to reproduce. If a woman is too thin, whether because of famine, self-starvation or too much physical exercise, she fails to ovulate and is rendered infertile [see "Fatness and Fertility," by Rose E. Frisch; SCIENTIFIC AMERICAN, March, 1988].

Since increased nutrition is a prerequisite for growth and since food is digested in the gastrointestinal tract, the stomach and intestines need to function optimally during periods of reproduction and intense growth. As the result of a complex reorchestration of the activity of the gastrointestinal hormones, they do just that. In my laboratory at the Karolinska Institute in Stockholm and in many other laboratories, the interacting roles of these hormones have been studied in pregnant women, in the fetus and newborn infant and in both the mother and child during lactation.

KERSTIN UVNÄS-MOBERG is senior lecturer in the department of pharmacology at the Karolinska Institute in Stockholm. She got her M.D. from the institute in 1970, her Ph.D. in 1976. She writes: "Having studied the physiology of the gastrointestinal endocrine system for 10 years and having gone through four pregnancies, I found it irresistible to try to put together the two areas of experience and knowledge."

To appreciate the role of the gastrointestinal tract in growth and reproduction, one needs first to consider the normal functions of the gut hormones. The hormones are polypeptides: short protein chains of from 10 to 100 amino acids. They are synthesized in specialized endocrine cells whose brushlike projections, the microvilli, project into the stomach and the small intestine. After a meal, different cell types sense either the distension of the wall of the gut or the presence of nutrients or of low *p*H (acidity) and are thereby triggered to secrete their hormones both directly into the gut and into the circulation. The activity of the gut's endocrine cells is also influenced by the autonomic nervous system. Activation of the vagal nerve of the parasympathetic nervous system promotes the release of the hormones that enhance digestion; activation of the splanchnic nerve of the sympathetic nervous system has the opposite effect.

The hormones in turn influence gastrointestinal motility and the secretion of digestive enzymes and acid. For example, the hormone gastrin, produced in the lower part of the stomach, enhances digestion by stimulating both gastric motility and secretion of gastric acid. The chemically related hormone cholecystokinin, secreted by cells in the upper part of the small intestine, inhibits the movement of food out of the stomach, thereby enhancing digestion and the absorption of nutrients into the circulation; it also stimulates the release of bile from storage in the gallbladder and the secretion of pancreatic enzymes. The hormone secretin, produced in the same area of the intestine, stimulates secretion by the pancreas of bicarbonate (which serves to neutralize gastric acid). Recently it has been established that these polypeptides also stimulate the growth of the organs they affect—in particular a thickening of the mucosa, the organs' mucous lining. In other words, they act in the gut as growth hormones.

One gastrointestinal hormone that does not assume these digestion-enhancing roles is somatostatin. This polypeptide was originally identified in the brain, in the hypothalamus. Now it is known also to be produced in the gastrointestinal tract, where particularly large numbers of somatostatin-

© 1989 SCIENTIFIC AMERICAN, INC

producing cells are present in the stomach and in the upper part of the small intestine.

Somatostatin exerts profound inhibitory effects in the gut. For example, it tends to decrease gastrointestinal motility and to block the secretion of hydrochloric acid in the stomach, the discharge of bile from storage in the gallbladder, the absorption of nutrient molecules through the intestinal wall and even the release of such hormones as gastrin and cholecystokinin. Somatostatin also counteracts the growth-promoting effect of gastrin and cholecystokinin in the gut. Because somatostatin is an inhibitory hormone, its neural control is opposite to that of the stimulatory hormones: its release is inhibited by vagal activity and stimulated by splanchnic activity.

Once absorbed, ingested food can be metabolized in essentially two ways. It can go the way of anabolism, the building up from small molecules of larger molecules that can contribute to growth or be stored in the body for future use. Or it can be broken down, in the process called catabolism, to provide energy. When the need for energy is high, as it is during physical exercise and psychological stress, energy is mobilized by catabolism in the liver, muscles and fat tissues in response to the activity of the sympathetic nervous system and the adrenal gland.

In contrast, anabolic metabolism dominates after ingestion of a meal, when fuel for the future is being deposited in the liver and in fat tissues. This storage of nutrients is promoted by the pancreatic hormone insulin, released when blood glucose levels rise in response to digestion of a meal. Eating also increases the activity of the parasympathetic nervous system and thus of the endocrine system of the gut. Several gut hormones, including gastrin, cholecystokinin and secretin, enhance the glucose-induced release of insulin, further stimulating the anabolic type of metabolism. Somatostatin, on the other hand, appears to inhibit the uptake of nutrients and their deposition in storage.

The fact that gut hormones exert an important influence on metabolism is nicely illustrated by some evolutionary data. Sture Falkmer here at the Karolinska has shown that the pancreas (which produces insulin) derives from the gut. In primitive vertebrates, which lack a separate pancreas, insulin is produced by endocrine cells in the gut and is released into the circulation after direct contact between those cells and the intestinal contents.

The gut hormones have psychological as well as metabolic effects. In times of stress, which favors the catabolic pathway of metabolism, the level of wakefulness is enhanced. After digestion, on the other hand, when nutrients are being stored, a person often is sleepy and has a sense of well-being. These postfeeding psychic effects originate to some extent in the gastrointestinal tract. Cholecystokinin injected into rats inhibits food intake; Robert S. Mansbach of the University of Vermont in Burlington and his colleagues found that in rats injections of the hormone may even cause electroencephalographic patterns characteristic of sleep. Cutting of the vagal nerve abolishes the inhibitory effect of cholecystokinin on food intake, suggesting that (at least in the rat) information moves from the gut to the brain along neural pathways.

The normal activity of the gastrointestinal tract changes in pregnancy, and the change is an important factor in the characteristic weight gain. Studies conducted in maternity-care centers in Sweden show that women put on an average of 15 kilograms during pregnancy. Part of

NURSING MOTHER AND CHILD are depicted in *Mary and the Child,* a late 15th-century Flemish work. The painting, by the artist known as the Master of the St. Catherine Legend, is in the National Museum in Stockholm. It now appears that breast-feeding promotes a hormone-mediated physiological symbiosis of mother and child.

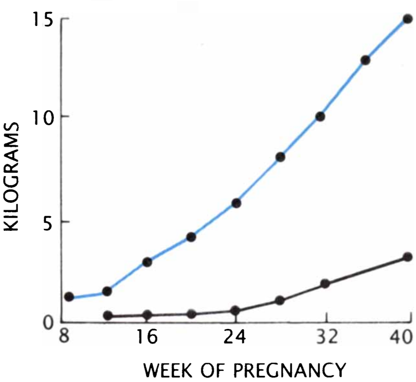
© 1989 SCIENTIFIC AMERICAN, INC

the gain reflects the weight of the fetus itself, the growth of the uterus and an increase in blood volume, but at least four kilograms represents the deposition of fat.

The simplest explanation for the weight gain of pregnancy would of course be increased food intake. (As folk wisdom has it, "She's eating for two.") Actually, increased intake is only partly responsible. As I mentioned above, women put on weight as early as during the first trimester of pregnancy, when most of them feel sick and consequently eat less, not more. Other physiological mechanisms must therefore be involved, and it is now clear that changes in the hormone activity of the gastrointestinal tract play a leading role.

My group and others have been able to show that the postprandial release of cholecystokinin increases in pregnant women. By working with pregnant dogs we were able to make repeated measurements that tracked a rise in the average plasma level of cholecystokinin. The elevation is maximal in the first third of pregnancy; from then on it declines gradually but remains clearly elevated above normal until delivery.

The changed hormone levels have several consequences. First, both the increase in cholecystokinin and the decrease in inhibitory somatostatin work to optimize the digestive process. Second, anabolic metabolism is facilitated and weight gain promoted, because the levels of the hormones that potentiate the glucose-induced release of insulin rise and the somatostatin level drops. Finally, the enhanced postprandial rise in cholecystokinin is probably responsible—by way of vagal signals to the brain—for the sleepiness and tiredness, particularly after meals, characteristic of early pregnancy. This fatigue has an adaptive effect: it tends to reduce physical activity, so that energy is saved and can be stored. (Of course many working women in advanced industrial societies keep working without letup almost throughout their pregnancy. Physiology suggests that this may actually not be the best course.)



PREGNANT WOMEN begin to store fat and gain weight (*colored curve*) in the first trimester, when they eat relatively little and before the weight of the fetus (*black curve*) is appreciable. The data are recent averages for Swedish women.

Early pregnancy is characterized not only by fatigue but also by such bothersome symptoms as intense hunger, intermingled with nausea; low blood pressure; and vertigo. The relation of the gastrointestinal tract to such symptoms is clear. Both the frequent hunger and the vertigo may be related to a fall in the blood glucose level, caused in part by the insulin-releasing effect of cholecystokinin and other gut hormones. The sickness and unpleasant sensations in the stomach are likely to be the result of delayed gastric emptying brought about by the high levels of cholecystokinin.

What brings about the changes in the gastrointestinal endocrine system during pregnancy? Hyperactivity of the vagal nerve (which, as I have indicated, modulates gut-hormone release) is likely to be one trigger. Koji Takeuchi of the Kyoto College of Pharmacy in Japan has shown that cutting the vagal nerve of pregnant rats sharply reduces the usual hypersecretion of such digestive juices as hydrochloric acid.

Why should that be? Vagal activity is affected by, among other things, the small neuropeptide oxytocin, which is produced by nuclei in the hypothalamus. Some oxytocin fibers project to the pituitary, from which the peptide is secreted into the circulation; other fibers project to the vagal motor nucleus in the brain stem and stimulate the vagal nerve [*see illustration at lower left on opposite page*]. Oxytocin secretion is powerfully enhanced by the steroid hormone estrogen, whose level rises in pregnancy. Gut-hormone release may also be affected locally by estrogen and another steroid hormone, progesterone.

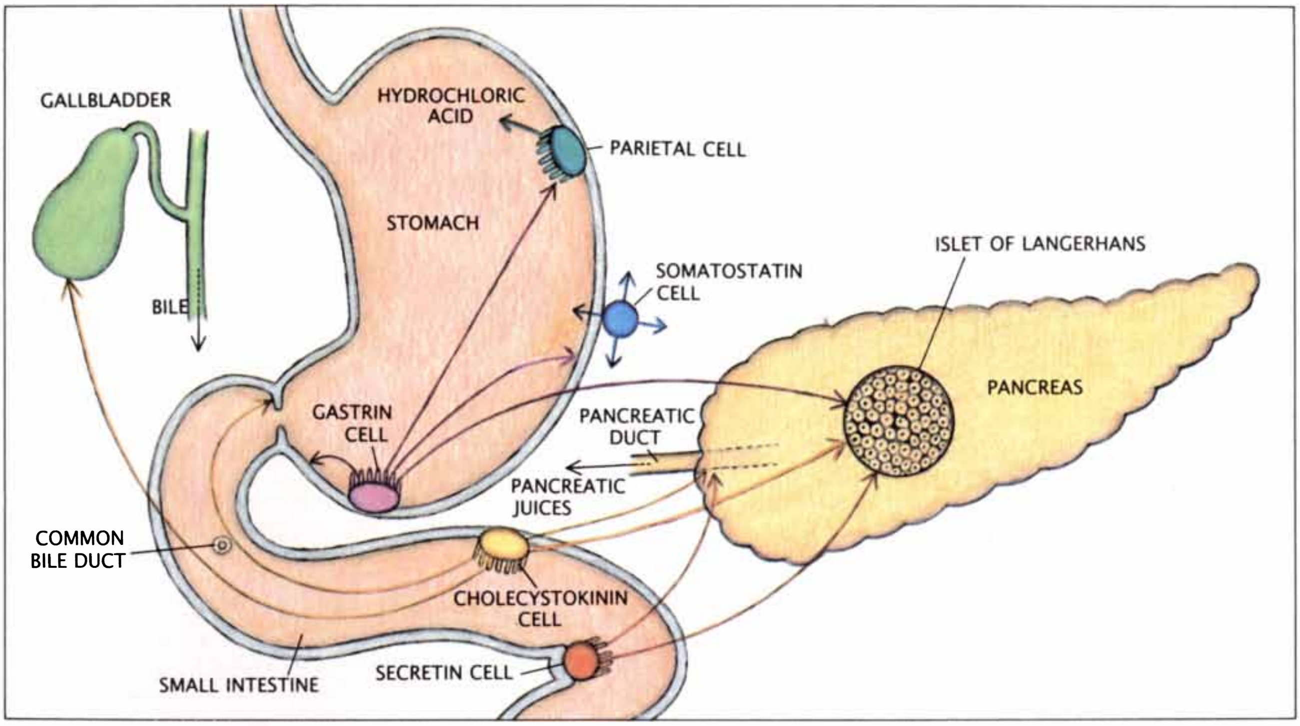
We modern human beings have inherited the genetic material and the physiology of our ancestors. For millennia any physiological mechanisms that made more energy available during pregnancy must have been critically important for the gestation and development of healthy children when food was sparse. In today's advanced industrial societies, characterized (for most people) by an abundance of food, the inherent ability of the female to cut down on energy expenditure during pregnancy is still being expressed—but it tends to be experienced only negatively, as unpleasant pregnancy symptoms and the risk of overweight. In a broader sense, both the ease with which most women put on weight and the higher frequency of obesity in women than in men may reflect women's latent and life-giving ability to store energy, an ability that is fully expressed during periods of active reproduction.

A woman's weight gain in pregnancy is far exceeded by the remarkable rate of growth of infants, who tend to double their birth weight within six months. Their food intake is proportionately high. A six-week-old baby weighing about four kilograms (less than nine pounds) drinks about 650 milliliters (22 ounces) of milk a day; for a 65-kilogram (143-pound) adult the corresponding volume of milk would be 10 liters! The average energy intake of calories in infants is four times as high per kilogram of body weight as that in adults. The gastrointestinal tract needs to be relatively large and highly active in order to cope with the amount of food and calories ingested in infancy. Because most of the nutrients are committed to anabolic metabolism for growth, the endocrine system promoting insulin release can be expected to be particularly active.

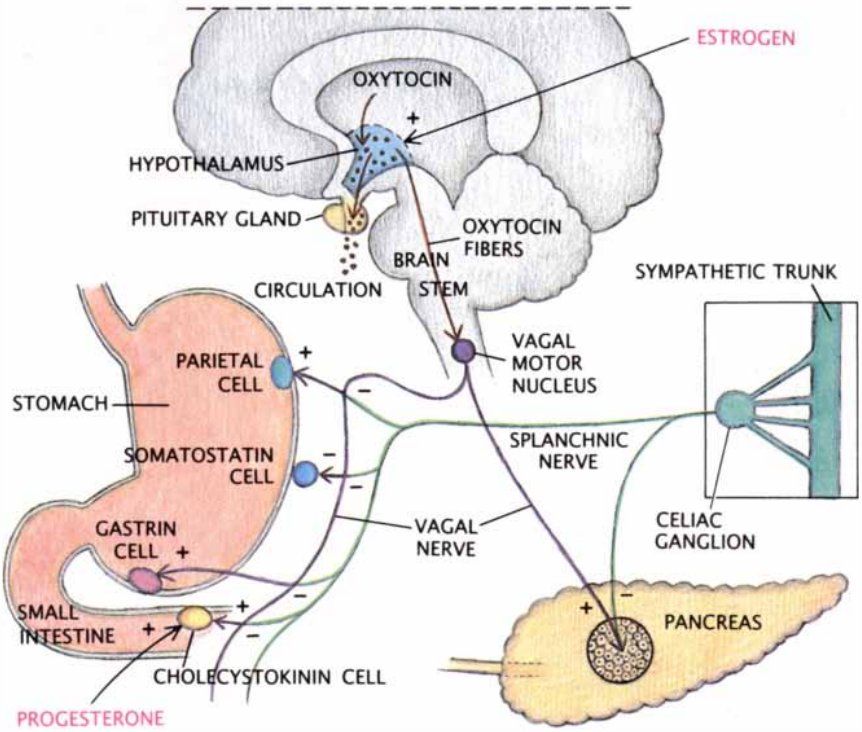
In support of this assumption, gastrin levels have been shown (by Alan Lucas of the Medical Research Council Dunn Nutrition Unit in Cambridge and A. Aynsley-Green of the University of Oxford and by my group) to be from five to 10 times higher in newborns than in grown-ups. These high gastrin levels in newborn infants are not secondary to a large intake of food. The baby actually ingests very little milk during the first few days of breast-feeding, and the high gastrin levels precede the high food intake that comes later in infancy.

The very early high gastrin levels can be explained at least in part by the fact that the newborn's gastrointestinal function has been prestimulated during fetal life. Even though nourishment is transferred to the fetus passively, by way of the umbilical cord, the fetal gastrointestinal tract is "trained" in utero for its future task of digesting food. It is known that from time to time the fetus swallows amniotic fluid; swallowing movements have been observed with ultrasound instruments as early as during the first

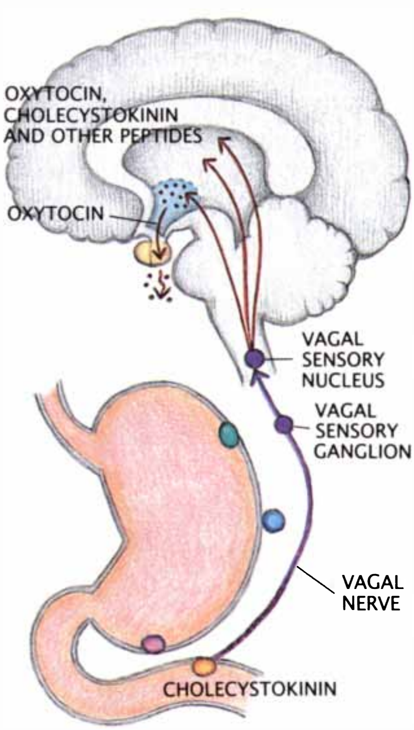
© 1989 SCIENTIFIC AMERICAN, INC



GASTROINTESTINAL HORMONES are secreted into the gut and the circulation by endocrine cells in the wall of the stomach and small intestine. Gastrin, cholecystokinin and secretin enhance pancreatic insulin secretion. Gastrin also stimulates release of hydrochloric acid, growth of mucosal cells and gastric motility. Cholecystokinin slows emptying of the stomach and stimulates discharge of bile from the gallbladder and secretion of digestive enzymes by the pancreas; secretin stimulates pancreatic bicarbonate secretion. Somatostatin inhibits secretion of gut hormones and counteracts their effects.



NEURAL CONTROL of gut-hormone release is exerted by the vagal and splanchnic nerves. Vagal fibers, activated in part by oxytocin, stimulate (+) or inhibit (−) endocrine-cell hormone secretion and pancreatic insulin release; splanchnic-nerve impulses counteract these vagal effects. The higher estrogen levels of pregnancy stimulate oxytocin secretion, enhancing the vagal effects on gut-hormone and insulin secretion; progesterone seems to act locally to enhance cholecystokinin release.



NEURAL SIGNALS also go from gut to brain. Cholecystokinin in the small intestine triggers vagal sensory impulses; these in turn activate vagal fibers that innervate the brain, promoting maternal behavior, satiety and sleepiness.

© 1989 SCIENTIFIC AMERICAN, INC

trimester of pregnancy by Heinz F. R. Prechtl of the State University at Groningen in the Netherlands. The amniotic fluid contains several substances, including epidermal growth factor and gastrin, that stimulate gastrointestinal maturation.

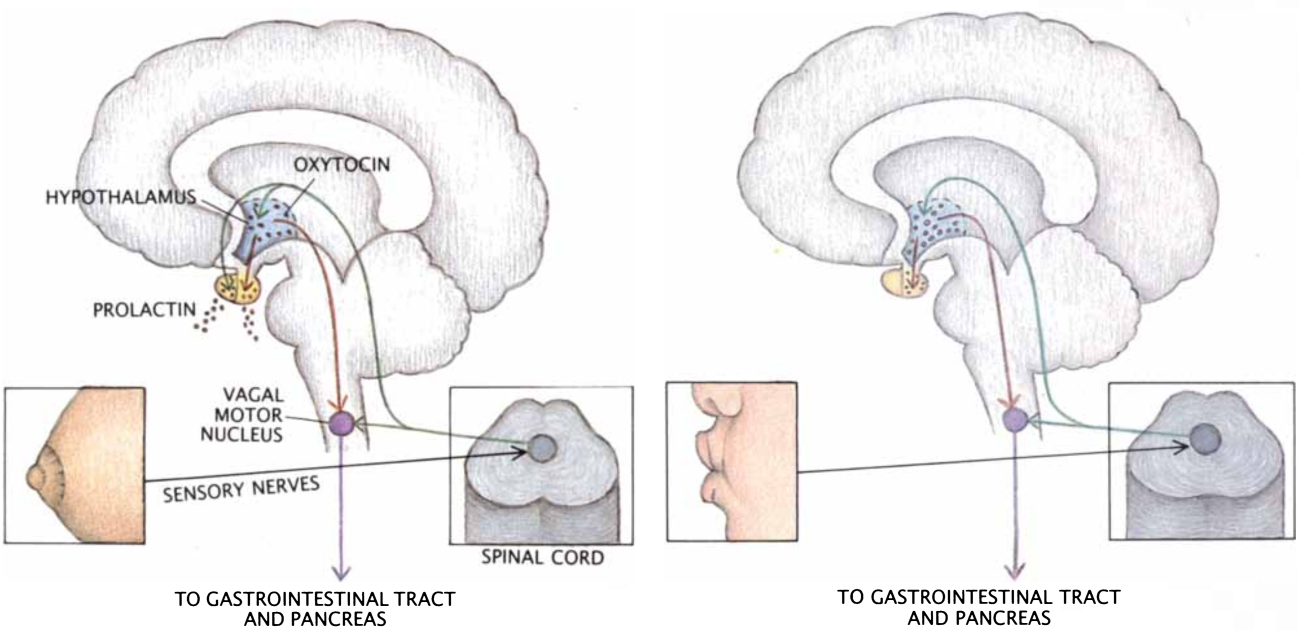
The importance of amniotic-fluid ingestion for the development of the fetal gastrointestinal tract was demonstrated by Sean J. Mulvihill of the University of California at Los Angeles School of Medicine. When he ligated the esophagus of fetal rabbits, the development of the gut was severely impaired; normal development was restored when bovine amniotic fluid was introduced into the gut. Ann-Marie Widström, Jan Winberg and I examined the stomach contents of infants immediately after birth. From our data we could conclude that in the human fetus periods of ingestion of amniotic fluid are followed by the secretion of fetal gastrin, somatostatin and gastric acid: apparently even in utero these substances are released in a time pattern resembling what is seen after a meal following birth.

A second reason for the high gastrin levels seen soon after birth may be that the newborn's suckling, by stimulating sensory nerves in the infant's mouth, triggers an activation of the vagal nerves and a consequent release of gastrin and other gut hormones. Giovanna Marchini and I tracked the suckling-related output of gastrointestinal peptides by monitoring their blood levels in plasma. When babies breast-feed, their gastrin, cholecystokinin and insulin levels rise. There is a first peak that is apparently mediated by the vagal nerve: it is notable that the levels rise similarly when babies merely suck on a pacifier. A second, more protracted hormone response results from the presence of food in the stomach and intestine, but it is probably augmented by the suckling-induced vagal activity.

There is substantial evidence that the suckling process itself has both physiological and psychological effects. Judy C. Bernbaum of the University of Pennsylvania School of Medicine found that infants that need to be fed by way of a nasal catheter grow faster on the same amount of calories if they suck on a pacifier while being tube-fed. Both breast-feeding and sucking on a pacifier sedate a baby and make it sleepy, presumably because suckling increases the cholecystokinin level. Pacifiers, incidentally, are often denigrated as being unhygienic or "spoiling" a child. On the contrary, giving a young infant a pacifier may have real physiological value. After all, the original breast-feeding pattern in humans (which is still seen in some hunter-gatherer tribes) included several periods of suckling per hour, day and night. Perhaps the pacifier serves to compensate for the reduced suckling time that results from the rigid modern pattern.

When infants are sick (whatever the specific diagnosis), they grow slowly and have gastrointestinal symptoms such as gastric retention, constipation and even vomiting. Marchini, Winberg and I found that sick children have 10-fold higher somatostatin levels than do healthy infants. In periods of stress (including stress caused by disease), the sympathetic nervous system is activated, mobilizing energy reserves. At the same time the function of the gastrointestinal tract is inhibited, partly as the result of an increased release of somatostatin. Because somatostatin inhibits not only gastrointestinal motility and secretion but also the release of hormones that stimulate anabolic (energy-storing) metabolism, the peptide is likely to lie behind the retarded growth as well as the impaired gastrointestinal function. Once again there is a clear connection between growth and the gastrointestinal tract—in this case expressed in a negative way.

From a physiological point of view lactation is a continuation of pregnancy: the mother's endocrine and digestive systems continue to provide nourishment for the child. The major differences are that the mother now stores energy in a special depot—the breast—and that the child now receives its nourishment in the form of milk rather than from the passage of nutrients through the umbilical cord.



SUCKLING stimulates the vagal nerve in the mother (*left*), by way of sensory receptors on the nipple, and in the infant (*right*), by way of receptors in the mouth. Vagal signals alter gut-hormone levels in the mother and the child, coordinating their metabolisms. Suckling also increases maternal prolactin and oxytocin levels, promoting milk production and flow.

© 1989 SCIENTIFIC AMERICAN, INC

A woman's need for calories is even higher during lactation than in pregnancy; it has been calculated that a lactating woman should increase her intake of calories by 25 percent above her normal intake. How is she to do it? Again the simplest way would be by eating more as the result of an increase in appetite. That does happen. Indeed, it has been shown in rats that the stimulus of the pups' suckling enhances the mother's food intake.

Yet surplus food may not be available during lactation, and so there is a need for physiological mechanisms other than hunger to meet the energy needs of mother and child. One way the mother can decrease her dependence on food is to consume the four kilograms or so of fat stored during pregnancy. That fat tends to be deposited on the thighs and buttocks, and energy is normally rather poorly mobilized from fat at those sites. Per Björntorp of the Sahlgrenska Hospital in Gothenburg has found that its mobilization is facilitated in breast-feeding women by an elegant physiological mechanism: the activity of the enzyme lipoprotein lipase, which promotes fat storage, is reduced specifically in thigh and buttock fat during lactation!

The fact remains that not all lactating women have stored fat during pregnancy, and so one can expect there to be other physiological mechanisms that save energy in the mother rather than making more energy available. Recently several investigators have found a discrepancy between the calculated ideal increase in energy intake to support lactation (25 percent in women) and the actual average increase achieved in lactating rats (which lack fat stores) and in women who have not stored fat or who breast-feed for a long time. More specifically, P. J. Illingworth of the Ninewells Hospital and Medical School in Dundee, Scotland, and his collaborators have found that less heat is generated in resting striatal muscles (which normally emit heat after a meal) in lactating women than in nonlactating women. This energy saving may in part explain how lactating women reduce their energy expenditure.

Even given the energy-mobilization and energy-saving devices, a major factor in supplying the increased energy required for lactation, as for pregnancy, is an increase in the activity of the gastrointestinal endocrine system. The heightened activity is in part secondary to higher food intake but is also caused by the suckling of offspring. Angelica Lindén, Maud Eriksson, Kerstin Svennersten, Widström and I have found, in women as well as in rats, pigs, dogs and cows, that each period of suckling is followed by a release of gastrin, insulin and cholecystokinin and by a decrease in the somatostatin level. The change in the endocrine pattern is a reflex response to suckling, mediated by the vagal nerves: the effects disappear in lactating rats when the vagal nerve is cut. (Indeed, it is likely that the activity of the entire parasympathetic nervous system is increased in response to suckling. The increase would counteract the activity of the sympathetic nervous system, and so it could be the cause of the lessened heat production in muscles, a catabolic process regulated by the sympathetic system.)

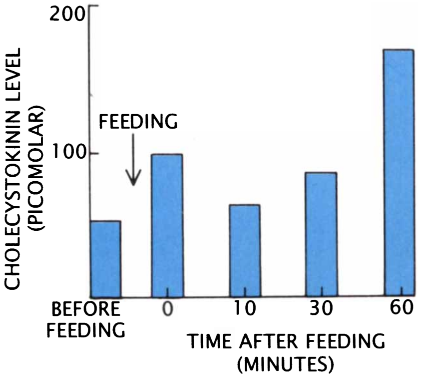
The physiological consequences of the release of gut hormones during suckling are significant. First, the digestive process is optimized, and the mucosa of the entire gut is thickened in order to meet the increased metabolic needs of lactation.

Second, energy is stored rather than being consumed for catabolic purposes. Energy is directed to a special depot, the mammary glands, by an increase in the pituitary hormone prolactin, which decreases the number of insulin receptors in maternal fat stores (thus reducing nutrient uptake there) and increases the number of receptors in the mammary glands, leading to an accumulation of nutrients there.

The suckling-related effect on the release of gut hormones contributes to an optimal balance between the intake and the expenditure of energy during lactation, since the amount of milk produced is regulated by the suckling stimulus. Bo Algers of the University for Veterinary Sciences in Skara and I found that suckling by piglets raises the maternal prolactin level and reduces the somatostatin level, and that the extent of these changes varies with the total amount of udder stimulation—its duration multiplied by the number of piglets. Another finding, by Widström and me, further demonstrates the importance of the gastrointestinal endocrine system for the success of lactation: the amount of the decrease in somatostatin level in lactating women that is caused by suckling is highly correlated with the amount of milk taken by the child.

Finally, suckling influences even the behavior of the mother. Women tend to feel sleepy when they are breast-feeding; rats exhibit electroencephalographic patterns characteristic of sleep when they nurse their young. Because cholecystokinin is released from the gastrointestinal tract in response to suckling, the peptide is likely to be involved, by way of the vagal nerves, in these psychological effects. The suckling-induced sedation may serve several purposes. For one thing, it saves energy; it may also help to keep the mother with her offspring.

It is of interest that, as the reader will recall, suckling by the infant triggers a vagally mediated gut-hormone response in the infant as well as in the mother. Indeed, the frequency and intensity of suckling regulate gastrointestinal function in both the mother and the infant and thereby synchronize their metabolisms. In other words, the mother and child are rendered symbiotic not only psychologically but also physiologically.



CHOLECYSTOKININ LEVEL in an infant is shown before breast-feeding, just after it and 10, 30 and 60 minutes later. The initial rise is a response to suckling; the level rises again when endocrine cells sense the passage of nutrients.

FURTHER READING

GASTROINTESTINAL HORMONES AND PATHOPHYSIOLOGY OF FUNCTIONAL GASTROINTESTINAL DISORDERS. Kerstin Uvnäs-Moberg in *Scandinavian Journal of Gastroenterology*, Vol. 22, Supplement 128, pages 138-146; 1987.

MATERNAL BEHAVIOR. Michael Numan in *The Physiology of Reproduction*, edited by E. Knobil et al. Raven Press, 1988.

NEUROENDOCRINE REGULATION OF HUNGER AND SATIETY. K. Uvnäs-Moberg in *Obesity in Europe*, Vol. 1, edited by P. Björntorp and S. Rössner. John Libbey & Company, Ltd., in press.

ROLE FOR SENSORY STIMULATION IN ENERGY ECONOMY OF MOTHER AND INFANT WITH PARTICULAR REGARD TO THE GASTROINTESTINAL ENDOCRINE SYSTEM. K. Uvnäs-Moberg and J. Winberg in *Textbook of Gastroenterology and Nutrition in Infancy*, second edition, edited by E. Lebenthal. Raven Press, in press.

© 1989 SCIENTIFIC AMERICAN, INC

Space Coloristics

The earth's surface presents an irreproducible palette to observers in orbit. Accurate reporting of its colors can reveal new facets of both nature and human vision

by Vladimir V. Vasyutin and Artur A. Tishchenko

For a cosmonaut, the strongest impression from space is that of the earth's bright and colorful landscapes and of the sunsets and sunrises—sights that cannot be reproduced or simulated by any technology. Photographs and other mechanical means convey neither the wealth of colors nor some of the complex surface patterns observed by the unaided human eye.

Describing the uncommonly bright and exuberant colors of natural formations and phenomena has for many years been an integral part of the Soviet cosmonaut program. From the time of the first space flights we have studied the ways in which the characteristics of color vision differ from what is common on the ground. We have also made use of important information about the earth that can only be gained by human observers. Space coloristics is the term given to the investigation of human color vision during space flight, the development of instruments for measuring it, the development and improvement of visual and automatic remote-sensing colorimeters, the study of the color attributes of natural objects and phenomena and the study of radiation spectrums and errors in the perception of spectral characteristics.

VLADIMIR V. VASYUTIN and ARTUR A. TISHCHENKO collaborate on the study of colors seen from orbit. Vasyutin is a pilot-cosmonaut currently working at the Gagarin Air Force Academy in Moscow. He graduated from the Higher Military Aviation Flying School in Kharkov in 1973 and became a cosmonaut in 1976. In 1986 he spent 65 days on board the *Salyut 7* space station, during which time he carried out coloristic experiments. He is a Hero of the Soviet Union. Tishchenko is the head of a laboratory at the Priroda (Nature) State Center in Moscow and chairman of the natural environment section of the U.S.S.R. Federation of Cosmonautics. He has helped develop visual, automatic and remote-sensing colorimeters for use in space. The coloristic experiments on board the *Salyut 5, 6* and *7* stations and the *Mir* orbital station were carried out under his supervision. This article first appeared in *V Mire Nauki,* the Russian-language edition of SCIENTIFIC AMERICAN.

Research began with the advent of manned orbital stations in the early 1970's. The stations made long missions feasible and permitted detailed investigations into human color vision and the perception of visual information from space. Even before then, physicians specializing in aviation and space medicine had obtained data on visual analysis during short space flights. Experiments during the Voskhod and Soyuz series of missions (1965–1969) had shown that during the first 24 hours of flight the cosmonauts' eyesight became less sharp as their bodies adjusted to microgravity, and their color vision underwent interesting changes. For example, the subjective perception of color brightness declined by 20 to 25 percent, especially at the red end of the spectrum.

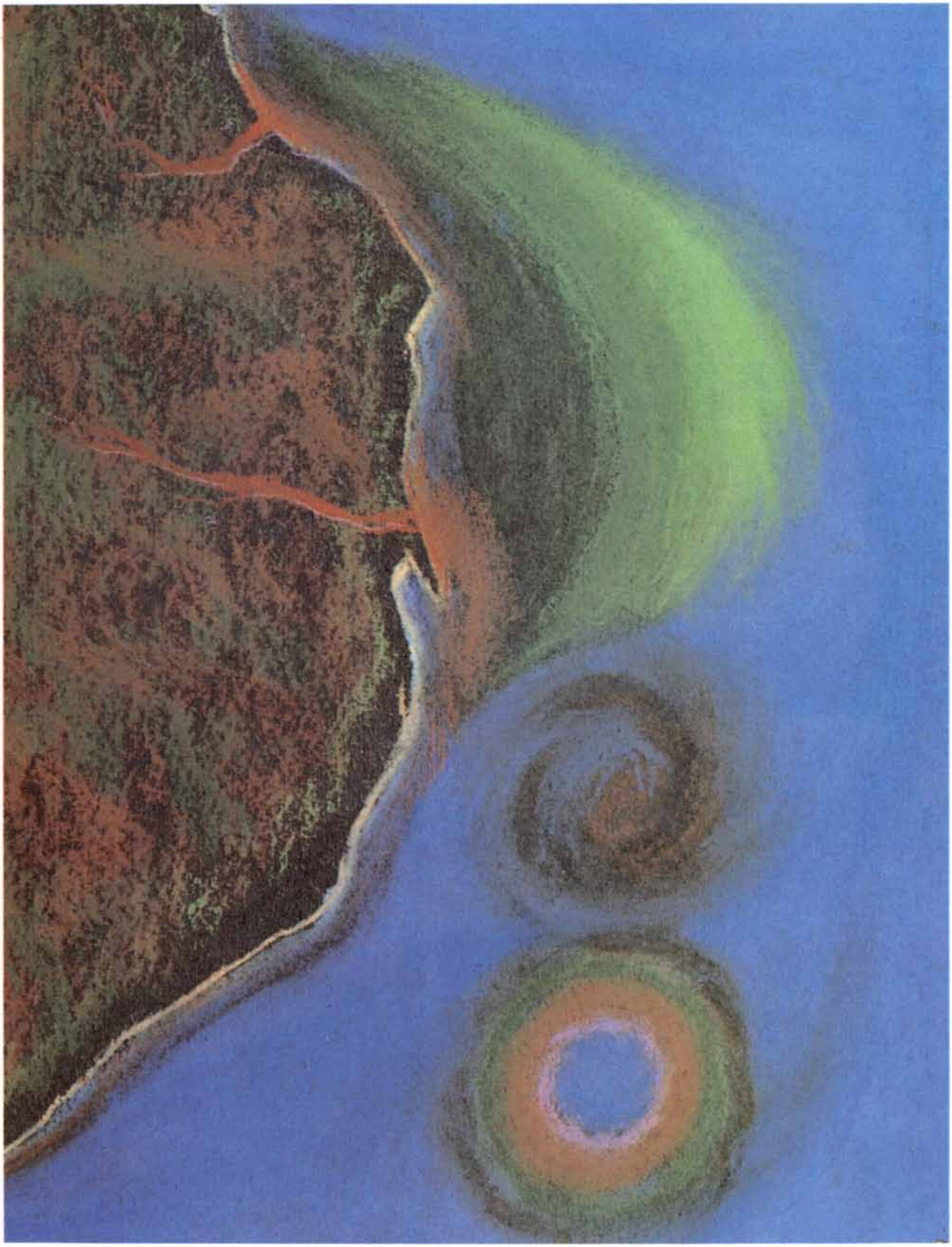
Even so, cosmonauts' eyes were still remarkable tools for observing features on the earth below; the human eye and brain have a number of properties not shared by mechanical remote-sensing systems. The eye's most surprising facility is its ability to detect and recognize complex patterns even when they are presented against a cluttering backdrop. This property is described as the constancy of eyesight. There are many kinds of constancy, such as that of the relative depth and orientation of objects. The most significant constancy is of color perception that prevails even when the light spectrum changes. This property makes it possible for observers to overcome obscuring phenomena such as atmospheric haze, blotting shadows and patches of sunlight. It is the principal advantage of remote-sensing systems based on the human eye.

Numerous factors enter into making good visual observations from space; some of them are subtle, such as the time taken for each observation. From orbit, the human eye can readily distinguish highways, ships, streets, forests, clearings and other surface features. One of us (Vasyutin) has noted, however, that the identification of the detail in a landscape and of relatively small objects depends largely on the duration of observation. The eye takes a few seconds to focus on small details. If the observer shifts his view to a different region, the eye defocuses. For example, it takes between three and five seconds to begin to distinguish a system of soil-erosion ravines in farmland. Cosmonauts who spend different amounts of time focusing on a feature before reporting their observations may "see" different things. The methodology for observing such formations has therefore been modified to allow for this effect.

Other changes in technique have become necessary over the years as well. One of the problems underlying visual observations in space, such as colorimetric observations, is training cosmonauts to overcome their color stereotypes, that is, an ingrained indifference to color information. It is also necessary to educate them so that they can sort out their subconscious color sensations into components such as hue, color and brightness, even under the complex conditions of space flight. Anatoli N. Berezovoi, who performed coloristic experiments on board the *Salyut 7* orbital station in 1982, was the first person to master this art. Not everyone has the knack; it requires both natural ability and training.

Someone whose eyesight does not have sufficient spectral resolution cannot identify objects with the varying surface structures characteristic of vegetation, water, deserts and so on. The gamut of colors in nature is re-

© 1989 SCIENTIFIC AMERICAN, INC



SOUTHERN TIP OF MADAGASCAR (shown with the east coast at the top) displays strong color contrasts in this drawing based on observations made while in orbit. The bright-green color near the coast shows plankton. Prevailing currents in the region create a vortex (circular formation) that mixes water from different layers in the ocean; the varying levels of suspended minerals give rise to contrasting colors. Plankton grow at the borders between different water masses.

© 1989 SCIENTIFIC AMERICAN, INC

nowned for its variety. A cosmonaut's eyesight must therefore meet the highest standards. Whereas a driver needs to distinguish only among the three colors of traffic lights, a cosmonaut must at the very least be able to distinguish the principal hues within a color, such as the different shades of brown corresponding to the different colors of sand in deserts.

The observation conditions on board a spacecraft are quite unusual. The illumination of an earthscape by the sun and the earth's scintillating atmosphere is complex and strange. Even experienced aircraft pilots who go into space speak of the very unusual nature of earthscapes and their unique colors. In addition, the speed of the craft in relation to the earth constantly changes the phase of an object's illumination and the angle from which it is viewed. Not only does the sky change color according to the weather and the angle of the sun, but the fine hues of the sky carry information about the atmosphere's purity and moisture content. Observation is often complicated by clouds that are too bright and by veiling atmospheric haze. The cosmonaut has to be able to distinguish these parameters from their color manifestations.

In addition to the requirements of good vision and training, observation of the earth from space depends on a number of technical factors. The most obvious one is the orientation of the spacecraft during the observational period: the craft must pass over (or nearly over) the area to be observed, and the portholes must face in the appropriate direction. The cosmonaut has very little time to observe a particular part of the globe. In a matter of 40 to 60 seconds, while the spacecraft is approaching the target area, the observer must locate a target by recognizing characteristic landmarks, prepare the recording equipment and aim it at the target, and then note the exact time of each photograph. This information is needed so that later on ballistic data on the spacecraft's orbit can be used to establish the angle, altitude and distance of observation as well as the angle at which the sun has illuminated the object.

The station portholes must also fulfill exacting optical requirements. They must not distort either the image passing through them or the color spectrum of the incoming light. This is particularly important because cosmonauts often have to observe low-contrast objects. Space-station portholes are made of a quartz glass that absorbs little visible light but keeps out short-wavelength ultraviolet light that could damage the eye. Portholes are protected by special covers that are opened only for observations.

Once the cosmonaut has made observations of objects on the earth below, those observations must be transmitted in a form that permits others to reconstruct what the cosmonaut has seen and to make use of it. Next to the camera, which transmits color information directly (but cannot make use of the pattern- and color-recognition skills of the human viewer), the next important tool has been the color atlas. The atlas is a set of reference samples with which observed colors can be compared. With the advent of the color atlas on board spacecraft and orbital stations in the late 1960's, cosmonauts switched over from impressionistic descriptions of colors based on associations or emotional content to a numerical coding system based on attributes such as hue, intensity and brilliance. Numerical coding has eliminated the uncertainties that previously characterized human descriptions of objects seen from space. It has also eliminated homespun descriptions such as "I see a lake with the color of a samovar," or "Just passed over the Sulak River delta, the alluviums are brick-colored," or "I can see mouse-colored plowed land." With a color atlas it suffices merely to relay

COLOR ATLAS AND COLORIMETER enable observers to determine the precise hue, saturation and brightness of natural features. The ATS-1000 atlas, developed at the All-Union Mendeleev Research Institute of Metrology, contains 1,000 standard samples, with a difference in wavelength between hues of about five nanometers. The atlas has been largely superseded by the Tsvet-1 colorimeter, which displays both the color standards and the object in question in a camera's viewfinder.

© 1989 SCIENTIFIC AMERICAN, INC

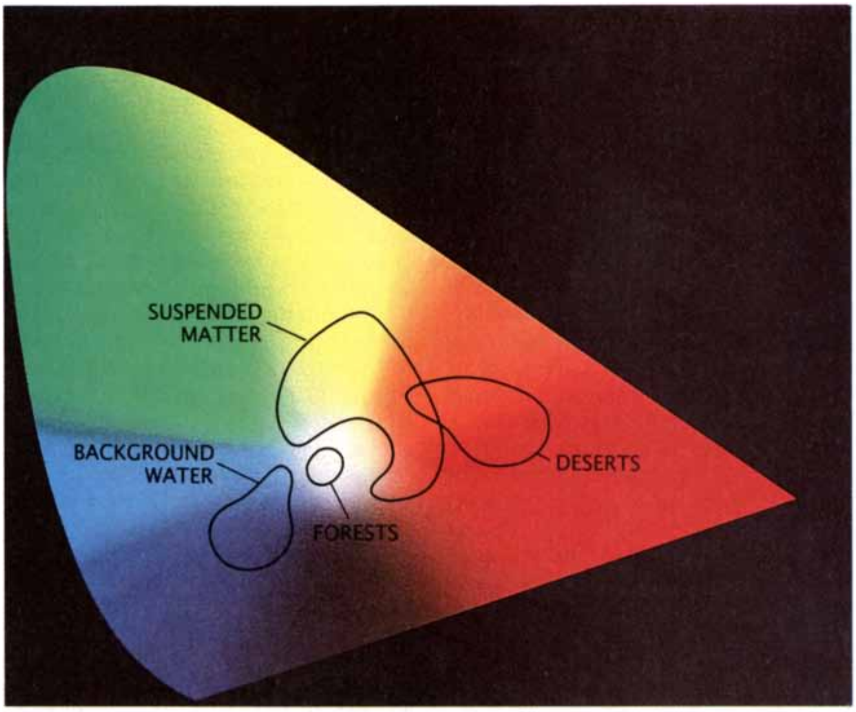
the color numbers back to the earth where the observations can be reconstructed using a similar atlas.

The first color atlas, with 30 hues, was used in 1968 by Georgi T. Beregovoi on board the *Soyuz 3* to measure the colors of dawns in space. The color intensity and luminosity of the sunrises, however, were such that the small atlas was not sufficient for describing them. In 1974 cosmonauts on board the *Salyut 4* orbital station used a sea-color atlas developed by oceanographers to match the Forel scale, which is based on a standardized set of 23 chemical solutions with colors ranging from blue to brown. In 1977 the *Salyut 6* cosmonauts tested an atlas of about 300 hues.

The experiences gained using these earlier color atlases in space demonstrated that a much greater diversity of colors was required to describe objects accurately. Nature could not be encompassed by such a limited number of samples. In 1982 cosmonauts on board the *Salyut 7* orbital station took with them the metrologically certified ATS-1000 atlas, developed in 1970 by Ye N. Yustova of the All-Union Mendeleev Research Institute of Metrology. The results in this article were nearly all obtained with this atlas.

The ATS-1000 color atlas consists of 1,000 samples whose hue, intensity and brightness have been measured precisely. In terms of hue, the accuracy of the samples is about five nanometers, which is close to the threshold value for color discrimination. Despite its imperfections—it is bulky and somewhat difficult to use at a porthole—and the subsequent development of an entire series of remote-sensing colorimeters, the ATS-1000 remains an invaluable tool; it often helps in unexpected cases. For example, the atlas was important to the success of the "electrotopographer" experiment carried out on board the *Salyut 7* by Aleksandr P. Aleksandrov and Vladimir A. Lyakhov in 1983. The experiment was designed to measure changes in construction materials exposed to space conditions; a special surface coating made visible microscopic surface defects on the samples. After the materials had been passed through the air lock, the cosmonauts noticed that the coating, initially grayish, had changed color, potentially skewing the experimental results.

The designers of the experiment and one of us (Tishchenko) came up with a technique for monitoring color by both the ATS-1000 atlas and spectral and luminosity measurements of the samples. The cosmonauts were able to correct the progress of the experiment and substantially increase the information obtained.



NATURAL FORMATIONS fall into a fairly narrow range of colors, as is shown on this color triangle. The different color regions are based on observations from orbit. Although each object on the earth has a unique color, its class can be determined swiftly from the range of hue, saturation and brightness into which it falls.

Between 1980 and 1983 scientists at the Priroda (Nature) State Center in Moscow developed the Tsvet (Color)-1, a portable visual colorimeter based on a single-lens-reflex camera. The colorimeter contains a removable cassette with color standards that can be seen through the device's viewfinder. The operator lines up a natural object in the viewfinder and then manipulates three handles to select a color standard matching the hue, intensity and brilliance of the color field. He then photographs the object and records the color readings in the observation log book. These data are relayed back to the earth during the next communication session. The Tsvet-1 was first used in space by Lyakhov and Aleksandrov on board the *Salyut 7* orbital station in 1983. The visual colorimeter achieves a greater accuracy of measurement than does the atlas. The atlas is also difficult to work with at a porthole because the light streaming through strikes the atlas so that a cosmonaut cannot compare hues effectively. With the Tsvet-1 the color standards are illuminated by an independent, stable and calibrated source of light.

The same period also saw the development of the Tsvet-2 automatic photoelectric colorimeter. The Tsvet-2 was utilized in research conducted in airplanes whose flights were timed to coincide with observations from orbit. Those who tested the device initially employed it to study the color of the ocean. The ocean appeared to be the simplest test of colorimetric methods because it is a vast object, covering most of the earth's surface, and changes little during the course of a flight. After conducting a number of experiments on board aircraft, the researchers were somewhat puzzled to find that the ocean presented an infinite variety of colors: its surface looked different at different times even in the same area.

The earliest systematic visual-instrumental observations of the earth were carried out on board the *Salyut 6* by Georgi M. Grechko, Vladimir V. Kovalyonok and Valeri V. Ryumin, the first very-long-term cosmonauts. Such observations had previously been impossible because it takes two to three weeks for a person's eyesight to adapt to conditions in orbit and for the person to open his eyes to the finer details of a landscape. They observed huge solitary waves 100 kilometers long, the traces left on the water surface by typhoons (which give an accu-

© 1989 SCIENTIFIC AMERICAN, INC

rate idea of the storms' size) and some of the characteristic features of the ocean floor. The ocean is eternally riveting with its constant mutability and its bright, exuberant colors.

In 1984 Leonid D. Kizim, Vladimir A. Solovyov and Oleg Yu. Atkov, the third crew of the *Salyut 7*, investigated the color of the Pacific Ocean; in a matter of two weeks they mapped its colors. Mapping by surface vessels would have taken between five and 10 years. Furthermore, surface measurements, which rely on the Forel scale of color solutions and a white disk lowered overboard, would have yielded only a limited number of color gradations.

The first systematic observations of the ocean from space were carried out by Kovalyonok in 1978 during his second mission on board the *Salyut 6* space station. He invited oceanographers to take part in these investigations—an important collaboration given the economic benefits of navigation and fisheries. The results of this cooperation were soon felt: ships began confirming high fish concentrations at locations indicated by cosmonauts; other ships received recommendations from space on how best to bypass ice fields and areas of great turbulence. (Since the 1940's fishing fleets have employed search planes carrying observers of uncanny eyesight; experienced "prospectors" can determine the kind of fish in a shoal from the shoal's shape and color.)

Kovalyonok's orbital observations and postflight evaluations of water color using an atlas provided the basis for the first true coloristic experiments in space. Berezovoi and Valentin V. Lebedev carried out the studies on board the *Salyut 7* in 1982. The two cosmonauts were specially trained in the theory and practice of coloristics. Their practical training took place on board a laboratory airplane on flights over the Black Sea, Caspian Sea, the Sea of Okhotsk and the Sea of Japan.

A month after the launch Berezovoi drew in his diary his first sketches of the surface structure of the ocean. During his 211 days in orbit he recorded more than 20 surface structures characteristic of the vertical and horizontal streams in the ocean, various water circulation patterns and formations associated with concentrations of organic and inorganic suspensions in the ocean depths. Choppy seas and wind currents were superposed on these structures, many of which had never been observed before.

In the course of this mission the cosmonauts made about 150 color measurements of the ocean. They recorded surface formations ranging from 10 to 500 kilometers in size. The observations identified three substantial areas against dark-blue backgrounds: light-blue "desert" waters, green areas rich in plankton and brown waters rich in suspended minerals and old plankton. The color contrasts between the objects observed and the background waters averaged about 10 nanometers.

The observations of the ocean off the coast of Argentina are a typical example of Berezovoi's colorimetric experiments. He investigated the color contrast of ocean currents, which have different temperatures and salt concentrations than the surrounding waters. The structures of the currents and their evolution over time are important subjects for research by oceanographers. Ocean currents significantly influence the temperature balance between the ocean and the atmosphere; they determine the weather both close to and far away from the currents themselves.

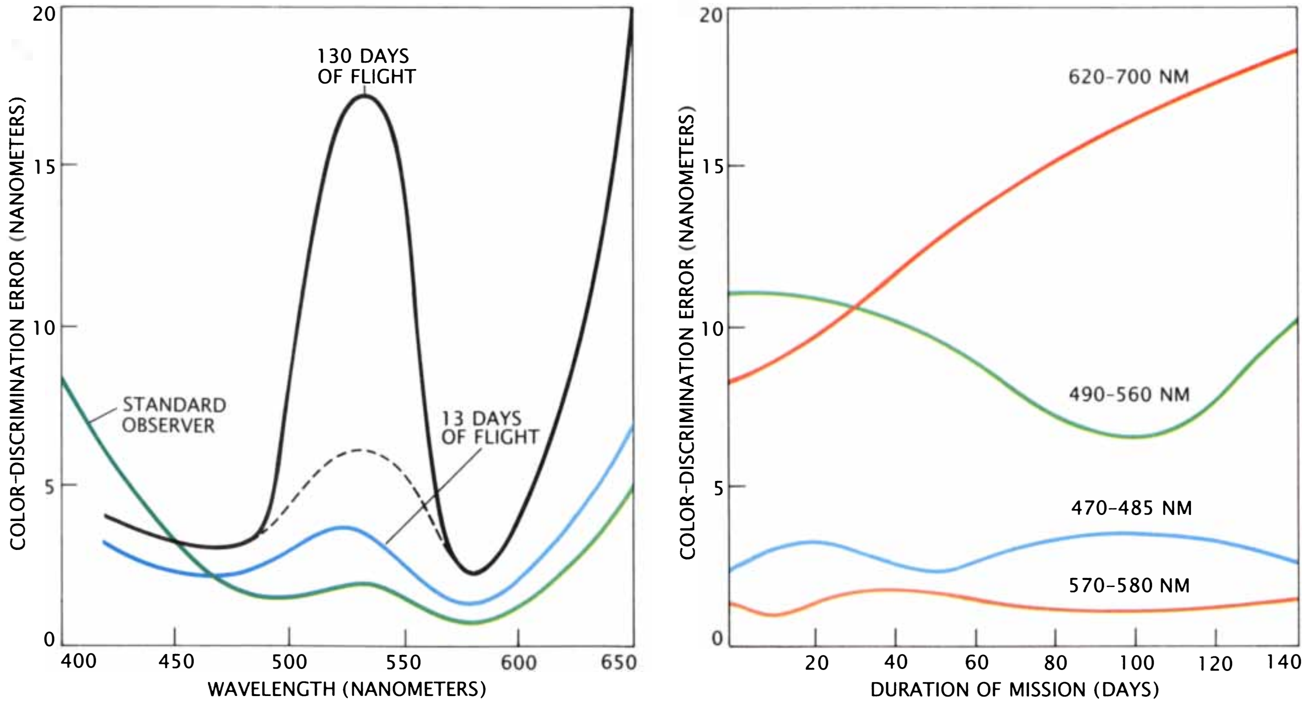
In addition to studying the features of the ocean surface, researchers also assessed the time required to observe those features from orbit. They recorded the time it took for a target object to be detected on approaching it, for the color attributes to appear as the spacecraft approached a position over the target and the time during which colors of the object and background could be recorded while flying directly over it.

The results were quite unexpected. When the color contrast was good—that is to say when the sun was about 30 to 45 degrees above the horizon and to one side or behind the observer—an object on the ocean surface could be observed for three to four minutes. Cosmonauts could observe high-contrast objects on land, such as oases in a desert, for up to six minutes—practically from horizon to horizon. Photographic and video recording, on the other hand, are only effective within about 45 degrees of the nadir, for an observation time of 1.5 to 2 minutes. The ability of the eyes to overcome the effects of atmospheric hindrances and observe features for a long time is the principal advantage of human observation.

ERUPTION OF ALAID VOLCANO on Atlasova island in the Kuril Islands, photographed from orbit, shows the detail observable from space. The almost imperceptible reddish hue of the smoke, whose precise color was reported by cosmonauts Vladimir V. Kovalyonok and Viktor P. Savinykh, provided volcanologists with important information on the chemical composition and consistency of the eruption.

In addition to gathering data about the features of the earth and its oceans, investigators have also studied the properties of human color vision during long space flights. This work was motivated primarily by the need to assess the accuracy of visual color measurements. Each cosmonaut carried out a series of tests in which he matched samples from a special table of 30 pigments (10 hues with three gradations of intensity and brightness for each) to the standard colors in a color atlas or a visual colorimeter. The difference between the perceived and actual hues determined the absolute error of the cosmonauts' color measurements. These tests were held repeatedly during long flights so as to build up sufficient data for

© 1989 SCIENTIFIC AMERICAN, INC



COLOR VISION OF COSMONAUTS differs significantly from that of typical ground-based subjects. Errors in determining the precise color of an object vary both with wavelength (*left*) and with the length of time spent in orbit (*right*). Cosmonauts show smaller errors in discrimination at the blue end of the spectrum and larger errors at the red end of the spectrum than do ground-based subjects. Their color discrimination in the blue-green and yellow-orange regions remains essentially unchanged. The error in discriminating green colors undergoes an unexplained cyclic variation in the course of long flights.

identifying regular patterns of change.

The tests yielded several patterns. The color discrimination at the blue end of the spectrum was stable, which means that visual observation of the oceans is effective. Color vision in the greenish-blue region (about 490 nanometers) and in the orange-yellow region (about 580 nanometers) of the spectrum was also quite stable, and cosmonauts' sensitivity to color differences was high. Color-discrimination errors were at a minimum. By a fortunate coincidence, greens, blues, oranges and yellows are the colors of most interest to life scientists, since they are the ones usually manifested by oceans, forests and deserts.

The cosmonauts performed most poorly on matching color samples at the red end of the spectrum. The error in discriminating red on the earth is quite high, but it gets even worse in space. Even so, cosmonauts were still able to identify the colors of large areas of low-intensity brown (as would be seen in some deserts and rocky areas) quite accurately.

The space experiments showed a cyclic pattern of error in the assessment of green: color errors increased for roughly the first four months of flight, then declined. Green is a critical color for gauging the overall quality of a cosmonaut's color vision because it appears so frequently in natural landscapes and because plants are of great biological and economic importance. Paradoxically, several cosmonauts have said after their missions that they did not see any green areas. Forests are very difficult to observe because the intensity of their green colors is at the eye's discrimination threshold. During one coloristic experiment Vasyutin was observing a forested area against the backdrop of the desert on the Mediterranean coast of Libya. The color coordinates he reported corresponded not to a green but rather to a grayish-lilac color. Subsequent careful observation of the forest showed that it had an earthy color but of very low intensity. By that time the cosmonauts had been in orbit for about a month.

The standardized tests of color discrimination present only part of the picture. In reality cosmonauts must observe natural objects under many different intensities and spectrums of illumination. The spectrums of reflected sunlight for different angles of the sun and atmospheric conditions have not been sufficiently studied. In many cases, however, cosmonauts are able to make finer discriminations of the colors of natural objects than the colors of those objects and the tests would suggest. This superior ability to discriminate the colors of real objects on the earth is not just a property of cosmonauts' eyesight but is also due to the enhanced color contrast of objects seen from space. All cosmonauts have commented on the brightness and intensity of the colors they saw from space. The reasons for this have yet to be clarified, although it may bear some resemblance to the effect of heightened color that artists achieve by putting a reflective varnish over the surface of a painting.

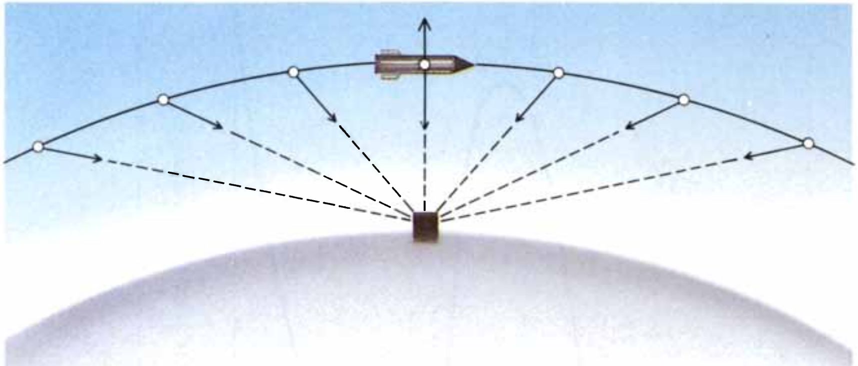
Some interesting data on this topic were obtained over the northern part of the Caspian Sea in the spring of 1982 during an experiment conducted simultaneously on board the *Salyut 7* station and an AH-30 laboratory airplane. Lebedev and Berezovoi, the *Salyut 7* crew, were to observe the vernal blooming of plankton in the Bay of Kizlyar and around the Agrakhansky peninsula. In mid-May the two reported intense green strips near the Agrakhansky peninsula and parallel to the coastline. On May 20 an airborne expedition took off for Astrakhan to make "below-the-satellite" observations. One of us (Tishchenko) was able to make a detailed study of the area and measure its colors. As seen from an airplane, the plankton grew in a

© 1989 SCIENTIFIC AMERICAN, INC

FIGHT HEART DISEASE, KIDNEY DISEASE AND BLINDNESS.

Support the American Diabetes Association.

Diabetes is a major contributor to heart disease, kidney disease and blindness. So when you support the American Diabetes Association, you fight some of the worst diseases of our time.



OBSERVATION CYCLE begins when an object appears over the horizon with respect to the orbital station. Depending on the object's form and contrast, it can take seconds or minutes to become visible. As the spacecraft passes overhead, seeing improves, and the colors of an object can be discerned. The process is repeated in reverse as the spacecraft moves away: first the color becomes indistinguishable, and then the object itself is lost in the background of the earth or the ocean. The exact length of time available for observation depends on the color contrast between the object and background and on the way the object reflects light. Objects on the ground reflect light diffusely, and ocean features produce mirrorlike reflections.

single 400-meter-wide strip consisting of many steps that waned in color intensity from the shore and lay at regular distances from the coastline.

Returning to Moscow on May 30, Tishchenko asked Lebedev and Berezovoi, "while on the subject," to specify the colors of the bright green steps. Surprisingly, they reported the colors of five strips, not one. The brightness and contrast of the strips as seen from space were much higher. This kind of color intensity can also sometimes be seen from an airplane over water from altitudes of about 10 kilometers in certain weather conditions. Cosmonauts on almost every subsequent mission have been asked to observe certain natural formations whose color is believed to be stable. The same landmarks have also been regularly studied from aircraft.

The coloristic investigations pursued thus far have provided the basis for developing new techniques of obtaining spectral information about the earth. Whereas techniques based on remote-sensing satellites such as *Landsat* measure the brightness of the ground below them in a number of spectral bands and then present this information in color-coded form, coloristic techniques based on human observation determine the characteristic colors of important natural features and present those colors directly. They also make use of the abilities of the human observer to perceive patterns that cannot be detected by photographic or other mechanical means. In view of their advantages, these remote-sensing techniques should hold considerable promise for both scientific and economic endeavors.

The investigations into the quality of color vision in space have also helped enhance the efficacy of cosmonauts' performance during tasks not involving the study of the earth. In addition, space coloristics has made it possible to formulate new visual standards for the successful study of nature from space that differ according to the type of object being observed. The first visual measurements of the color parameters of little-known natural objects have made it possible to draw up better specifications for remote-sensing spectrometers and photographic systems. The unusual conditions of living in space and the new factors—such as changes in vision—affecting life and work that have been identified in orbit may contribute to the realization of as yet dormant human potentials.

FURTHER READING

COLOR INVESTIGATIONS OF THE EARTH. A. A. Tishchenko in *Aviatsiya i kosmonavtika (Aviation and Cosmonautics)*, No. 5, page 46; 1983.

TSVET-1 COLORIMETER. A. A. Tishchenko in *Aviatsiya i kosmonavtika*, No. 1, page 44; 1984.

CHANGES IN COLOR VISION PARAMETERS IN LONG-TERM SPACE FLIGHT. V. V. Vasyutin, A. A. Volkov, A. A. Tishchenko and V. Ye. Fokin. (Collection of Scientific Readings on Cosmonautics, in Memory of Gagarin.) Nauka, Moscow, 1986.

SPACE COLORISTICS. A. A. Tishchenko in *Tekhnicheskaya estetika (Human Engineering)*, No. 5, pages 3–5; 1987.

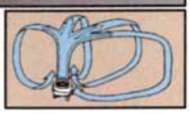
© 1989 SCIENTIFIC AMERICAN, INC

THE LIFESTYLE RESOURCE™

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

AERODYNAMIC COOLING

Turbo-Aire's jetstream



Hassock position–venturi effect

The Turbo-Aire fan moves more air and provides more cooling than ordinary fans - you get more than twice the air delivered by oscillating room fans. Patented and computer designed to maximize efficiency and minimize noise and vibration. This fan creates an exceptionally strong, smooth jetstream column that sets a room stirring with secondary air currents. Refreshes better than the intermittent blast of air from an oscillating fan. Aerodynamic housing increases blade-tip efficiency over conventional fans. Floor, table or wall mount, adjusts to any angle. Set in the hassock position, the five bladed, 12″ fan redistributes air in an entire room, making it as useful in winter as in summer. Can reduce air conditioning costs and is 300% more energy efficient than ordinary fans with the same type and size motor. With a 3 speed fan control, the Turbo-Aire is the most advanced 12″ portable house fan. Weighs less than a case of cold soft drinks - and more easily moved! Manufacturer's 2-year limited warranty. UL listed. **$79.95** #2161.

AN AMERICAN CLASSIC

The *very best* of the Adirondack Chairs are still made in the Adirondacks, in Willsboro, N.Y.. Willsboro's skilled woodworkers have done more than just re-create the original Adirondack Chair - they've added refinements such as a fully contoured seat and 6″ wide arms for the ultimate in comfort. And the workmanship! The seat and backrest are hand-dowled and glued to the frame for strength, using solid hardwood pegs (not capped screws) in the true Yankee manner. The rich rock maple is an extremely dense and attractively grained wood, all sanded - no sharp corners or splintering. A unique (and patented) tongue-in-groove folding mechanism lets you fold the chair to a mere 10″ thickness for storage. Each completed piece is treated with a wood protectant, ready for you to apply a clear finish or paint, or leave it to weather naturally. There has never been an Adirondack Chair like this...and we have a matching footrest, designed to create a chaise lounge and an end table to go along side - both constructed in the same fine tradition. The chair is 38″ x 14″ x 20,″ **$139.95** #1881. The footrest is 14″ x 20″ x 22,″ **$79.95** #1891. The end table is 16″ x 14″ x 24,″ **$69.95** #1901. Some assembly required.

ZONE OF CALM

Noise produces stress in the human organism. Today the volume of civilization nearly everywhere seriously interferes with our abilities to relax, read, sleep, concentrate or otherwise function at optimum efficiency. But you needn't be victimized by noise pollution. The *new* Marsona Sound Conditioner electrically synthesizes a variety of pleasing natural sounds that help mask out and reduce the annoyance of unwanted noise. A 5 inch speaker brings you the sounds of ocean surf, summer rain, mountain waterfalls — as close to nature as you will find. You control the volume, the wave pattern, wave or rain rhythm and the seeming nearness or distance of the source. Solid state electronics designed for continuous use. UL listed. Manufacturer's 1 year limited warranty. **$139.95** #2202.

WORRY-FREE WATERING

Water grows things best when it arrives at regular intervals, a fact that sometimes gets by Mother Nature. Now give your lawn or garden the advantage of timely watering. RainMatic 2000 automatically turns water on and off, up to four times a day. Set any length watering interval with your own schedule or use the simple built-in program. Manually override anytime. Great for sprinklers and trickle systems. An absolute necessity during restricted watering. Operates all season on 4-C batteries (not included). Weather-protected. LCD readout with digital clock. Locks on faucet. Manufacturer's 1-yr. warranty. **$49.95** #1081.

THE LIFESTYLE RESOURCE™

FOR FASTEST SERVICE — CREDIT CARD ORDERS CALL TOLL-FREE 24 HRS A DAY **800-872-5200**

DEPT. SFAG09; 921 EASTWIND DR. SUITE 114: WESTERVILLE, OH 43081

	ITEM NO.	QTY.	DESCRIPTION	ITEM PRICE	TOTAL PRICE
SEND TO (PLEASE PRINT)					
ADDRESS					
CITY STATE ZIP					
☐ CHECK OR MONEY ORDER ☐ MASTERCARD ☐ VISA ☐ AMEX SIGNATURE					
ACCT. # EXP. DATE				SUB TOTAL	
				SHIPPING (See table at left)	
				TOTAL	

Shipping Charge covers UPS, handling and insurance for guaranteed delivery.
UPS Second Day available for an additional $7.50 per order.

Up to $20....$ 3.95		$50.01 to $60....$ 7.95	
$20.01 to $30....$ 4.95		$60.01 to $70....$ 8.95	
$30.01 to $40....$ 5.95		$70.01 to $100....$10.95	
$40.01 to $50....$ 6.95		Over $100....$12.95	

UPS 2ND DAY AIR

Canadian residents please call (614) 794-2662 for ordering information.

ORDER WITH CONFIDENCE

- *Most orders ship within 48 hours of receipt.*
- *Credit Card orders billed only upon shipment.*
- *No risk 30-day return privilege.*

THE MEDIA DEVELOPMENT GROUP

©1989 C51 ADVERTISEMENT

© 1989 SCIENTIFIC AMERICAN, INC

Visual Thinking in the Ice Age

After some 2.5 million years in which the archaeological record reveals very little innovation, body ornamentation and visual images explode onto the scene in Europe about 35,000 years ago

by Randall White

For the first 2.5 million years of the archaeological record the only artifacts of human beings and their hominid precursors are strictly utilitarian: stone tools. Then, about 35,000 years ago in Europe, a dramatic turning point was reached. Along with new kinds of stone tools and implements made of bone and antler, the first objects of a symbolic nature appear. These include both the first adornments of the body, in the form of beads and pendants, and the first known attempts at rendering nature, both in two dimensions and in three. This cultural explosion, which marks the beginning of the period known as the Aurignacian, took place at the same time over large parts of western and eastern Europe. Why?

It has long been assumed that the answer lies in the realm of biology: anatomically modern human beings (*Homo sapiens sapiens*) evolved, replacing the preceding Neanderthals and bringing into play significant new neurological capacities. Recent work, however, makes that conclusion seem less than likely. It now seems that *Homo sapiens sapiens* evolved 100,000 years ago (or more) in Africa and moved into Europe several thousand years before the beginning of the Aurignacian. For most of that period our subspecies had nothing that could be called symbolic. Hence, the appearance of ornaments and representations seems to have been the result of cultural processes, not biological or neurological ones. The study of the first material images offers promising insights into this landmark of cultural evolution.

RANDALL WHITE is associate professor of anthropology at New York University. Born in Canada, he earned his Ph.D. in anthropology at the University of Toronto before obtaining a teaching position at N.Y.U. in 1981. Some of the work described in this article grew out of the recognition that American museums held massive collections of French Ice Age art and artifacts, most of which had never been studied or described in scholarly publications and were thought by French prehistorians to be lost. He has now analyzed more than 2,000 Ice Age ornaments and incomplete beads. In 1986 White was guest curator of an exhibit at the American Museum of Natural History on Ice Age art called "Dark Caves, Bright Visions."

The beginning of the Aurignacian period in Europe is generally thought to coincide with the larger transition from Middle Paleolithic to Upper Paleolithic culture. As long as the appearance of anatomically modern human beings was dated to about 35,000 years ago, this transition could be accounted for by the appearance of *Homo sapiens sapiens*. In recent years, however, skeletal remains of *Homo sapiens sapiens* that date to about 90,000 years ago have been found in the Near East and in South Africa. Studies of mitochondrial DNA suggest an origin in Africa for all modern human populations, perhaps as recently as from 50,000 to 250,000 years ago. That model, if true, implies that modern human populations all descend from groups that expanded out of Africa after that time.

The expansion seems to have been completed in western Europe by about 30,000 years ago, when the more robust and powerful Neanderthals were replaced by the more gracile (slender) modern types. This replacement appears to have been completed fairly quickly—within as little as 5,000 years. By the time it was over, the practice of ornamenting the body (now characteristic of all human societies) had appeared in full force in Europe and as far afield as Australia. The richest body of evidence, which comes from a group of Aurignacian sites in France, Belgium and Germany, is under intense investigation by me and other archaeologists. In this article I shall describe the earliest known body ornaments and the attempts to understand their relation to other developments in the transition to the Upper Paleolithic, including the appearance of the first representations of nature.

WOOLLY MAMMOTH was sculpted from mammoth ivory at Vogelherd in Germa-

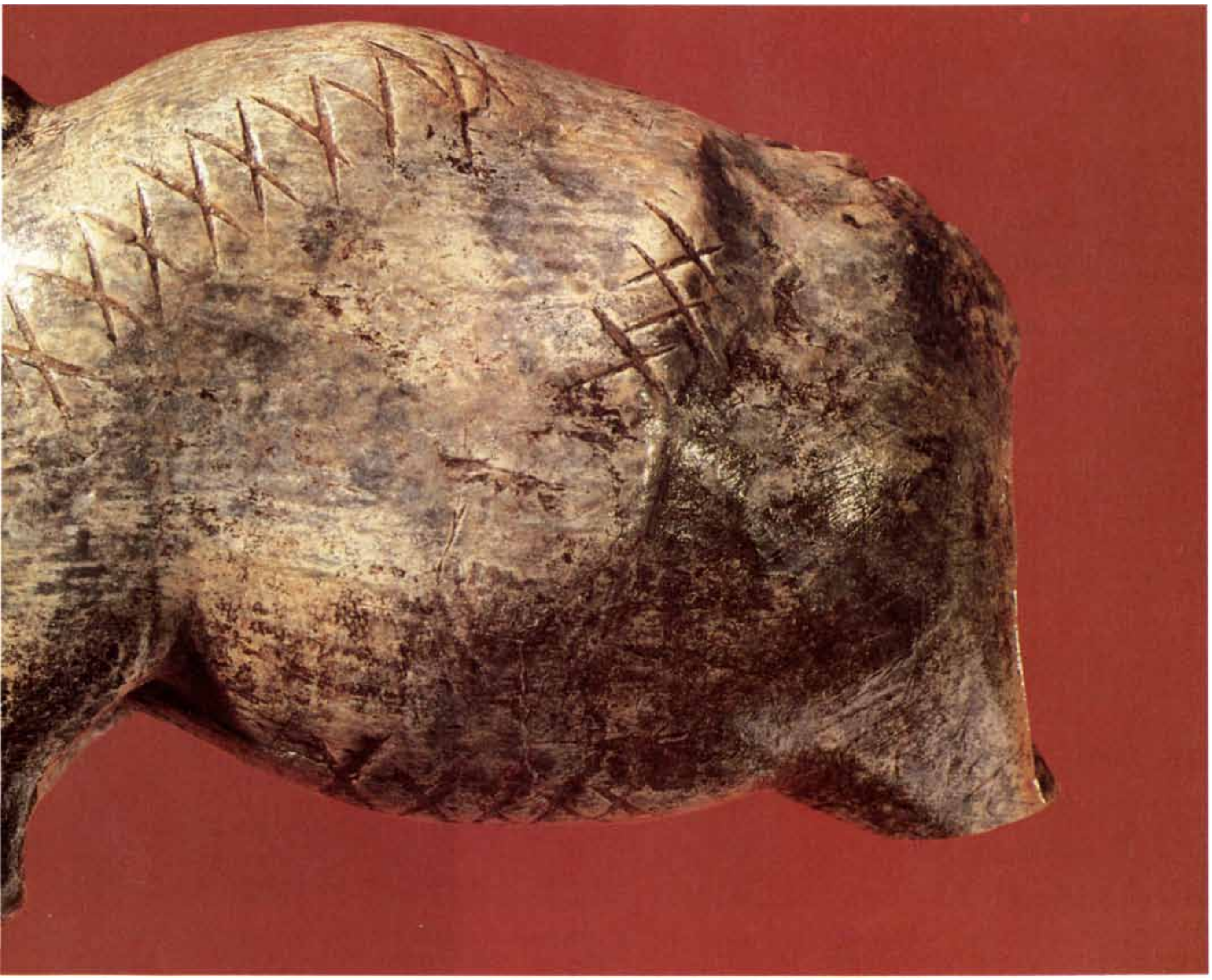
© 1989 SCIENTIFIC AMERICAN, INC

The greatest abundance of early body ornaments is associated with the first phases of the Aurignacian, lasting from about 34,000 to about 30,000 years ago. The same cultural layers from which the ornaments come also reveal well-developed industries in stone, bone and antler. The stone tools include tiny flint bladelets as well as implements for scraping and for engraving made by modifying high-quality flint blades. The bone and antler work includes awls, wedges, spear-points and smoothing tools. This assemblage of tools helped to sustain relatively small groups that hunted large herbivores such as the reindeer, woolly mammoth, wild horse, bison and European red deer. From the same period comes the first substantial evidence for riverine fishing and the hunting of birds.

At the hundreds of sites known to have been occupied by Aurignacian groups, archaeologists frequently find objects that they interpret as beads or pendants. These objects are made of a variety of materials, including soft stone, shell, tooth and, most frequently, ivory from the tusks of woolly mammoths. Many of them have been perforated for suspension, and it is known from the succeeding cultural period (called the Gravettian) that such beads and pendants were indeed used for decorating the body. At the site of Sungir, near Moscow, which has been dated to between 28,000 and 24,000 years ago, thousands of carefully shaped ivory beads and pierced teeth, assembled in long strands, were found covering the remains of several individuals.

Even earlier, in the late Aurignacian, ornaments were associated with burials. One of the best-known Aurignacian sites that includes body ornaments is the classic site of Cro-Magnon in the Vézère Valley of southwestern France. In 1868 railroad workers there discovered four adult skeletons and a fetal skeleton, along with numerous pierced shells and an ivory pendant. The assemblage dates to about 28,000 years ago, at the end of the Aurignacian.

Yet the very earliest body ornaments, from the beginning of the Aurignacian, have not been found in association with burials. Instead, the beads and pendants are found by the hundreds in the campsites and on the living surfaces occupied by early members of *Homo sapiens sapiens* in Europe. Robert H.



ny from 34,000 to 32,000 years ago during the Aurignacian cultural period. During the same period body ornaments and two-dimensional images appear in the archaeological record for the first time. The significance of the X pattern is not known.

© 1989 SCIENTIFIC AMERICAN, INC

Gargett of the University of California at Berkeley has recently called into question the idea that Neanderthals ever buried their dead [see "Science and the Citizen"; SCIENTIFIC AMERICAN, June, 1989]. If Gargett is correct, the first interments may not have taken place until well after *Homo sapiens sapiens* appeared in Europe. Whether he is correct or not, the earliest ornaments must be interpreted exclusive of the burial evidence. This puts limits on how far one can push the interpretative process, because without burials it is difficult to say much about the differences in body decoration that might have been associated with critical social variables such as age and gender.

Before turning to interpretation, however, I would like to give some sense of where the body ornaments are found—in terms of distribution at Aurignacian sites—and how they were made. Although the decorative objects are quite abundant overall, their frequency varies greatly from site to site. Some sites have few, if any, ornaments despite excellent conditions of preservation; others have yielded hundreds of ornaments. Many of the ornament-rich sites include not only finished objects but also scrap material and partly finished work. This spectrum of production materials indicates that the ornaments were in fact made at the site and, as we shall see, enables archaeologists to reconstruct the sequence of production steps.

One of the most striking things about Aurignacian ornaments is that they were manufactured primarily from materials exotic to the sites where they are found. During the Aurignacian the capacity of human groups to procure materials over great distances seems to have increased sharply. Certain types of soft stone, shell and ivory are found at sites hundreds of kilometers from where they were naturally available. It is not known whether such rare and exceptional materials were obtained by travel or through the development of exchange networks among different groups, but it is striking that in most cases such exotic substances were transformed into body ornaments and not into tools or weapons.

Such a pattern surely indicates that the people of the Aurignacian put great value on the "foreign" materials and that these materials were exploited for the purposes of social display. What is more, the largest proportion of the exotic materials are found at

BEAD PRODUCTION during the Aurignacian period in southwestern France has been reconstructed by the author. The starting point was a rod of mammoth ivory (*1*) that was incised around its circumference to form points for breaking. When broken, each rod yielded a series of bead "blanks" (*2*). The blanks were shaped to yield a thick "bulb" at one end and a thin "stem" at the other (*3, 4*). The blank was pierced (*5*), then heavily ground and polished, leaving a finely polished bead (*6*).

© 1989 SCIENTIFIC AMERICAN, INC

ornament-rich sites. Why some sites should be so rich in ornaments and exotic materials is not yet understood, but it is possible that the rich sites were prearranged places where the exchange of highly valued materials took place, perhaps attracting groups of people from faraway. There are other possible explanations, but it does seem probable that the ornament-rich sites represent special contexts where social display was of the utmost importance.

However the exotic materials became concentrated at particular sites, their transformation into ornaments seems to have been guided by a complex conceptual framework. The ornament makers were not gathering their raw material at random. On the contrary, they were making conscious choices from the range of available materials. For example, one large class of ornaments consists of pierced animal teeth, but when teeth were made into pendants, only those of certain species were selected. Generally these were carnivores. There are exceptions: the site of Mladeč in Czechoslovakia has yielded an abundance of pierced beaver teeth. Yet at most Aurignacian sites the preferred species were hunters such as the fox, the wolf, the hyena and the bear.

It seems quite likely that the ornament makers of the Aurignacian were, by choosing the teeth of certain species, attempting to evoke or assume some of the overall qualities of those species. The conceptual process whereby a part is made to stand for the whole is called metonymy and is well known from both literature and myth. Homer, for example, sometimes refers to ships as "sails." More to the point, a metonymical basis for body ornaments is well known to anthropologists from small-scale societies studied today, whose members have both material and ritual means of assuming the qualities of animals. The fact that carnivores were chosen is perfectly consistent with this explanation: as their midden debris (garbage) shows, the Aurignacian groups sustained themselves in large part by hunting.

If the conceptual basis for Aurignacian ornaments seems relatively complex, the technology exploited to realize those concepts varies from elaborate to quite simple. The simple end of the continuum is represented by pierced seashells and animal teeth. Shells and the roots of animal teeth are seldom thicker than .2 centimeter.

MIMICRY OF NATURE was one aspect of Aurignacian body decoration. The bead at the left is a pierced red-deer canine tooth. The one at the right is a limestone imitation. Both are from the site of Trou Magrite in Belgium and are about 30,000 years old.

The Aurignacians penetrated them by gouging from either side until it was possible to break through simply by applying pressure with a sharp object. Drilling was virtually never used, except for the softest substances. Indeed, early Aurignacian tool assemblages are notable for the lack of the fine flint drills commonly found at later sites. Another technique was to incise a groove around the root that probably served for attaching a cord.

Shells, along with certain types of teeth and bones, come in "perforation-ready" packages. Mammoth ivory, on the other hand, required more elaborate methods of preparation. When Aurignacian groups obtained pieces of mammoth ivory, they worked it down into units of acceptable size for their own form of production. The criterion of acceptability varied considerably among regions and was ultimately tied to regional differences in aesthetics and symbolism. The description that follows concentrates on southwestern France, because that is the region that has provided the largest sample of body ornaments—and also because that is where much of my own work has been centered.

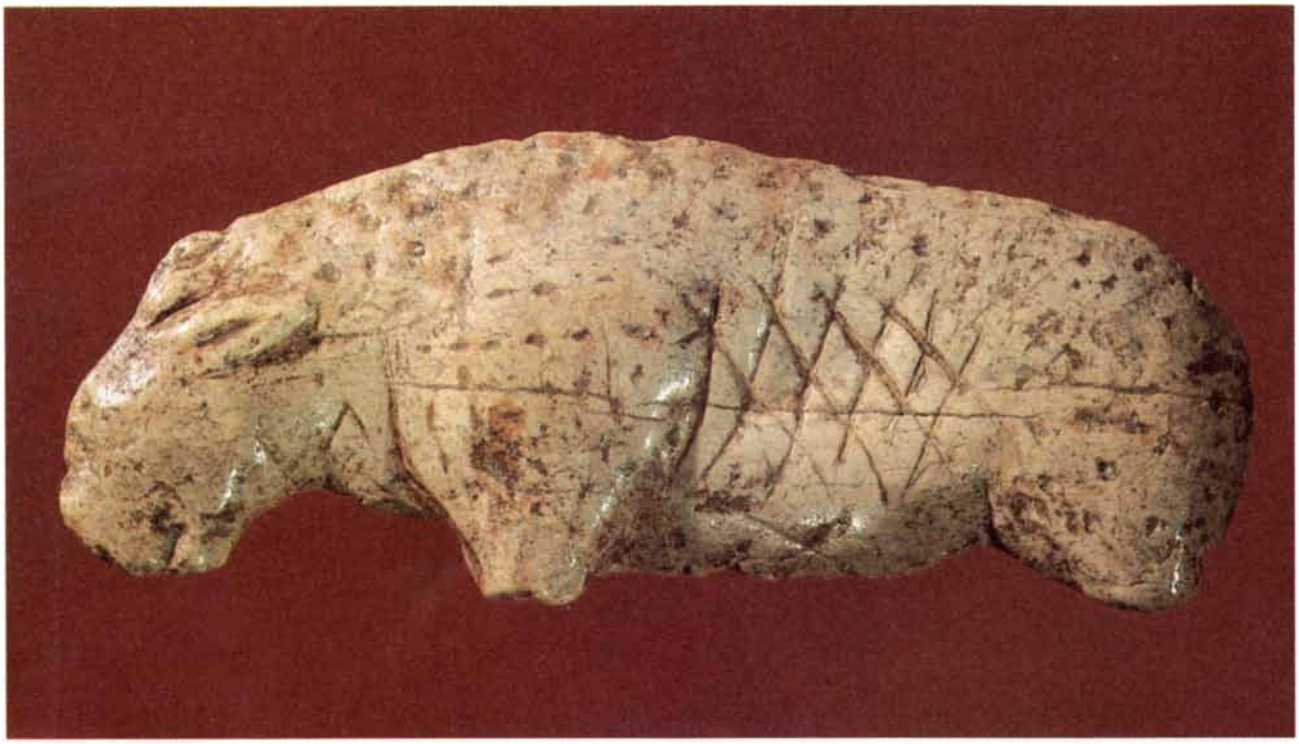
The production of ivory beads in southwestern France began with long "sticks" of mammoth ivory. These were apparently brought from some distance, since there are few discarded tusk fragments and no mammoth bones in the French sites. The ivory sticks had been removed lengthwise from the outer layer of the tusk, which is concentrically laminated. The laminated layer provided the best material for working with Aurignacian techniques; the unlaminated inner core, which is extremely hard, was either discarded or on occasion (at German sites) sculpted into animal figurines.

After the sticks of ivory had been obtained, they were whittled down into pencil-like rods. Each rod was then segmented by means of incisions around its circumference at intervals of one to two centimeters. The segments were then snapped off to form bead "blanks" for further production work; hundreds of these blanks have been identified. At the point where they were snapped off, the fine layers of the ivory always run perpendicular to the ultimate direction of the hole through the bead. The choice of direction was conscious and was made so that one end of the blank could be thinned by removing the outer laminae, first on one surface and then on the other.

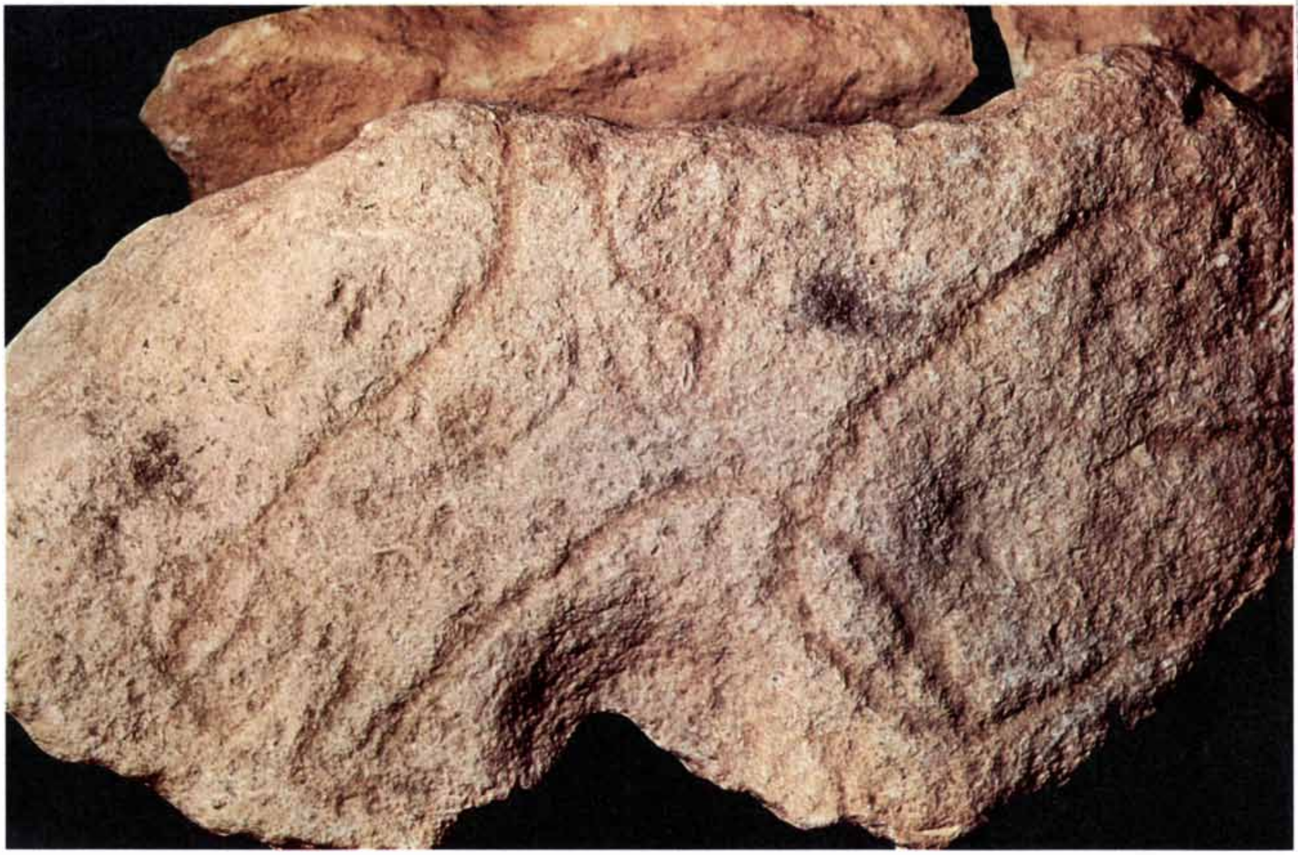
The result of this thinning was a roughed-out bead with a thin upper half ("stem") and a thicker, rounded lower half ("bulb"). At this point the blank bears a certain resemblance to certain types of animal teeth, particularly the vestigial canines of the red deer and reindeer, and this mimicry was, I believe, the aim of the process. (As we shall see, imitation of natural forms and designs seems to have been a hallmark of early Aurignacian representational processes.) The thinning of the blank almost always resulted in an upper portion .2 centimeter or less thick, which was then pierced. The hole was made where the stem and the bulb met.

After being pierced, the entire object

© 1989 SCIENTIFIC AMERICAN, INC



SCULPTED FIGURE found at Vogelherd in Germany is thought to be a feline, partly because of the shape of the clearly visible ear. The coat is indicated by punctuations. The sculpture, some 32,000 years old, is made from mammoth ivory.

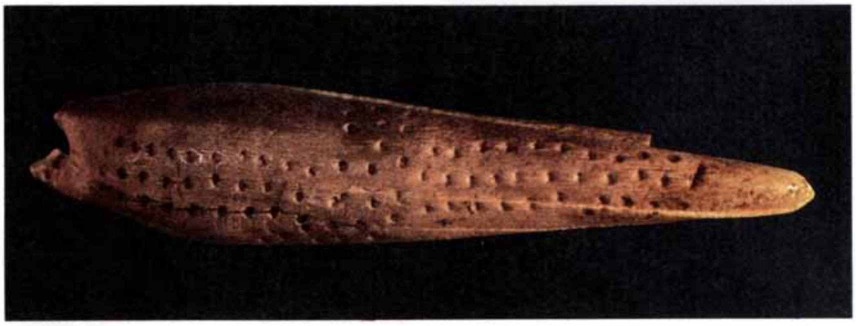


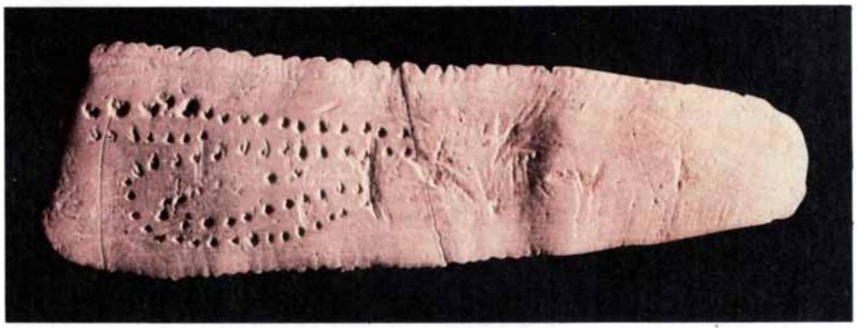
ENGRAVED FIGURE on a flat block of limestone is generally interpreted as a horse's head. The separate figure at the right appears more abstract. Found in 1927 at the site of Abri Cellier in France, the engraved block is about 32,000 years old.

© 1989 SCIENTIFIC AMERICAN, INC

was ground and finely polished. These procedures (not generally associated with Paleolithic technology) removed more than half of the original mass of ivory. In particular, the stem was almost entirely ground away and the bulb rounded and smoothed. The final product often prompts people to ask how the Aurignacians were able to make such delicate holes. In actuality, by first making the hole and then grinding the blank down around it, they were disguising their inability to do just that.

How were the products of this process worn? It is possible that they were assembled into necklaces or bracelets. Another possibility is that they were sewn onto clothing, forming designs or patterns. Careful examination of the beads, however, reveals that the majority of them are intact. If they had been individually sewn onto clothing, one would expect that they would have to be broken to be lost. Sometimes paired sets are found at the same site, which suggests earplugs or earrings. There are also togglelike beads that may have been buttons or, again, earplugs. Undoubtedly there were different methods of attachment, corresponding to the variety of artifacts, but the limiting reality is that we do not for the moment know how the earliest ornaments were worn.





MEANDERING PUNCTUATIONS are found on many Aurignacian objects. The objects shown here are about 32,000 years old. They come from Abri Blanchard in southwestern France. The one in the middle has been interpreted by some anthropologists as a fish or a seal. Alexander Marshack has argued that the marks on the plaque at the bottom constitute a lunar calendar. The presence of similar markings on other Aurignacian objects, however, suggests the intent was aesthetic rather than notational.

Nor can archaeologists know exactly what the sudden, almost explosive, appearance of these ornaments implies about social development during the Ice Age. Yet information from modern, small-scale societies provides some clues. In such societies ornaments help to define and communicate social identity. Gender, social role, economic status and group affiliation are all expressed through ornamenting and otherwise decorating (with tattoos, for example) the body. Indeed, Terence Turner of the University of Chicago has described the body in modern society as a "symbolic stage" on which the drama of socialization is enacted. This symbolic drama is a human universal that is found in all known societies.

If adornment is a vehicle for constructing and representing social identities, what does its initial appearance mean? There would seem to be two general answers to the question. The first is that the appearance of body ornaments is linked to the elaboration of the social categories themselves. In other words, it was during the Aurignacian that institutionalized differences in social status first became manifest. The second general possibility is that the early Aurignacian was the period when the techniques of representation first became adequate to the task of embodying the nascent social distinctions. These two explanations are not, of course, mutually exclusive: it is possible that new social identities developed simultaneously with the methods for expressing them.

Unfortunately the archaeological record—particularly in the absence of burials—is not sufficient to enable us to discriminate very finely among the various possibilities. Yet it is clear that the early Aurignacian was the setting for remarkable advances in techniques of representation. The Aurignacian sites richest in body ornaments are precisely the ones that contain some of the earliest known works of art in the form of engraved and painted limestone slabs, mammoth-ivory sculptures and objects of bone, antler, ivory and stone decorated with puzzling—but clearly intentional—imagery.

The earliest graphic images include representations of natural forms and other simple, redundant designs that seem more abstract. The representational images include both engravings and paintings of animals and engravings that have been taken to represent the female genitalia. Although many archaeologists find that interpretation unconvincing, it has gained some support from two clear phallic images: a sculpture in animal horn and a limestone engraving. Among the more abstract motifs are patterns of X's,

© 1989 SCIENTIFIC AMERICAN, INC

ANTHROPOMORPHIC FIGURE, probably depicting a human female, was carved from mammoth ivory. Only about five centimeters tall, the figure was found at Abri Cellier in France. The head and hairline are clearly visible. The paired marks are common on Aurignacian objects; their significance is not known.

notches, incisions and punctuations (gouged-out points). Such abstract images are found at site after site over broad regions, suggesting they were signs and symbols shared by members of regional social entities that may just have been coming into existence during the Aurignacian.

It is my hypothesis that even some of the more abstract-seeming designs may actually be representations of natural objects. One of the best-known decorated Aurignacian objects is an ivory plaque from the Abri Blanchard rock-shelter in southwestern France. One face of the plaque bears a series of punctuations that researcher Alexander Marshack has interpreted as an early lunar calendar. Many archaeologists doubt this interpretation, partly because the Abri Blanchard and adjacent sites have yielded many objects with similar designs that do not appear to have any calendrical function. It seems more reasonable (and just as significant in evolutionary terms) to argue that the aim was aesthetic, not notational.

To my knowledge, the possibility that the Abri Blanchard plaque is actually representational has never been considered. Yet findings from a nearby site, Abri de la Souquette, suggest that it may have been. At Abri de la Souquette five ivory pendants were found, covered with rows of punctuations. In examining them I was struck by their resemblance to the patterns seen on Atlantic seashells from the same site. In fact, the pendants were faithful replicas of the shells, which bore the same meandering rows of punctuations. It may be that many, if not all, of the punctuated objects represent a transfer of natural patterns to new contexts where they became forms of decoration.

The transfer of qualities from one context to another is an essential part of the construction of metaphors, and another object from Abri de la Souquette seems even more immediately metaphorical than the pendants. That object is a split-based spearpoint made from reindeer antler. Such points are common in Aurignacian sites, but this one has been modified in a curious way. The pointed end has been whittled down, giving the impression of a snout. An unusually fine hole pierces the object where an eye might be. The split base is not unlike a pair of fins. Overall, one gets the subtle (but powerful) impression of a seal.

If this interpretation is correct, the point is a material analogue, or metaphor, for the marine mammal. In it we see the transfer of animal attributes (snout, eye, fins) to an implement that initially had some formal similarities to the animal (mainly in the base, which resembles fins). Intriguingly, the remains of a seal were found at another early Aurignacian site only 100 meters away. La Souquette is 200 kilometers from the sea, and although seals have been known to swim far upriver, their presence there during the Aurignacian would no doubt have been a source of great curiosity.

The "seal" point (if that is what it is) is something of an exception at the French sites. The relatively few animal images there are almost always engravings in limestone and usually so crude they are difficult to identify. In contrast, three early Aurignacian sites in Germany (Vogelherd, Geissenklosterle and Hohlenstein-Stadel) have yielded some 20 ivory statuettes, often marked with abstract designs. Some of these are so well executed and realistic that scholars have assumed there must have been earlier prototypes that have not yet been discovered.

Those who look for prototypes tend to take for granted the notion that the earliest art must be simple and childlike. The archaeological record of the Aurignacian suggests that the earliest drawing is just that. The contemporaneous sculpture, however, is lifelike and expressive. Perhaps it was easier to reduce an animal to a three-dimensional scale model than it was to depict it in two dimensions. It seems to have taken thousands of years for early artists to develop the conventions and tricks of two-dimensional representation seen at sites such as Lascaux, which was painted 17,000 years after the first Upper Paleolithic images were created.

Having surveyed some of the key developments of the Aurignacian in technology, body ornamentation and representation, one is led to ask whether there is any way to tie them together. I think that the answer is yes, at least tentatively. Close to the heart of these developments is an increased ability to think in—and communicate by means of—specific visual images. This increased competence might help to explain the burst of technological progress that characterizes the Aurignacian. Heidi Knecht of New York University has recently argued that an emerging consideration of design possibilities by way of two- and three-dimensional images stimu-

© 1989 SCIENTIFIC AMERICAN, INC

lated the rapid development of new tools and weapons.

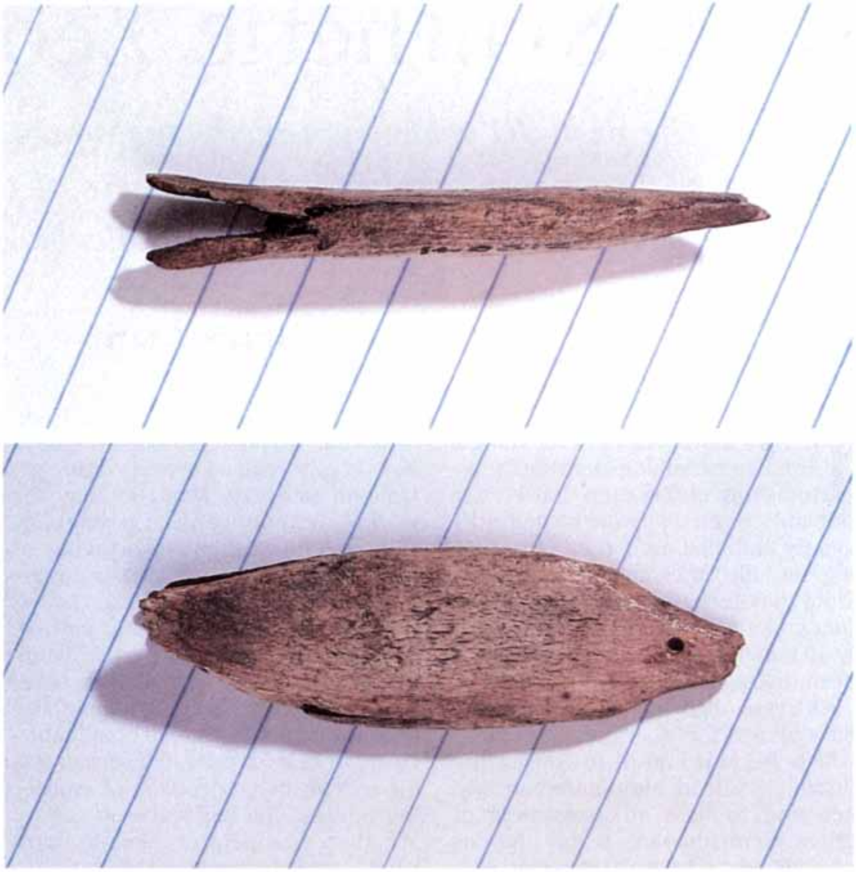
Innovations in the domain of spatial and visual imagination are not unrelated to body ornamentation. Ornamenting the body is a way of constructing and objectifying a set of social distinctions that are not otherwise visible. Socially defined categories of persons do not have obvious physical correlates unless they are endowed with such correlatives by human beings. By abstracting certain formal properties of natural objects, either metonymically or metaphorically, and bestowing them on certain classes of people, the Aurignacians were able to make manifest gradations of social status as well as overall group identities, for which there is no evidence prior to the Aurignacian.

This new ability to isolate attributes and then transfer them to another context would have had profound implications. It would have become possible to transfer properties such as "pointedness" or "barbedness" from natural contexts to technological ones. Having abstracted the concept of "barbedness" from its natural context, the Aurignacian toolmaker might then have been able to visualize alternative designs for a barbed spearpoint. This kind of visual thinking is now so common in technical contexts we have difficulty realizing that it may be a relatively recent and revolutionary cultural development.

Just how revolutionary a development visual thinking may have been is suggested by the work of Brooke Hindle, for many years a historian of science at New York University. Hindle's studies of the Industrial Revolution emphasize the significance of image manipulation. He points out that almost all the great industrial inventors thought in images, and he emphasizes the importance of the exploded drawing for technical innovation. Perhaps it is no coincidence that the world's first representational images appeared during one of the greatest periods of technological and social innovation in all of human history.

In some form or other behind all of these developments must stand language. To provide the base for further social and technical change, images must be shared and communicated. Language may well have existed before the Upper Paleolithic, perhaps in a concrete form closely tied to specific natural objects. The revolutionary developments of that period, however, presuppose innovations in the manipulation and sharing of images. Although certain types of neural "hardware" were no doubt a prerequisite for these innovations, the inception of image-based representation should not be seen as the crossing of a neurological threshold. On the contrary, it was more probably a cultural transition based on the establishment of shared conventions of representation.

The first realization of these possibilities was at least as significant evolutionarily as the first use of fire or stone tools. The consequences for invention would have been profound. Compared with the preceding Neanderthals, who displayed virtually no technological innovation during more than 100,000 years of existence, the Aurignacian people were able to realize with increasing rapidity a wide range of social, technological and mythical possibilities. Much of this rapid evolutionary development, which continues today, is no doubt due to the forming, manipulating and sharing of images.



SPLIT-BASE SPEARPOINT may represent a seal. The base, originally intended to fit into a wood shaft, may depict fins (*upper panel*). The tip has been shortened and pierced where the eye would be (*lower panel*). The spearpoint, made of reindeer antler, was found at Abri de la Souquette near Sergeac in southwestern France. Intriguingly, the remains of a seal were found at another Aurignacian site less than 100 meters away.

FURTHER READING

AURIGNACIAN SIGNS, PENDANTS AND ART OBJECTS IN CENTRAL AND EASTERN EUROPE. J. Hahn in *World Archaeology*, Vol. 3, No. 3, pages 252–266; Feb., 1972.

THE CREATIVE EXPLOSION. J. E. Pfeiffer. Cornell University Press, 1985.

DARK CAVES, BRIGHT VISIONS: LIFE IN ICE AGE EUROPE. Randall White. American Museum of Natural History/W. W. Norton & Company, 1986.

THE ORIGINS OF IMAGE MAKING. Whitney Davis in *Current Anthropology*, Vol. 27, No. 3, pages 193–216; June, 1986.

PRODUCTION COMPLEXITY AND STANDARDISATION IN EARLY AURIGNACIAN BEAD AND PENDANT MANUFACTURE: EVOLUTIONARY IMPLICATIONS. R. White in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*, edited by Christopher Stringer and Paul Mellars. Edinburgh University Press, in press.

TOWARD A CONTEXTUAL UNDERSTANDING OF THE EARLIEST BODY ORNAMENTS. R. White in *Patterns and Processes in Later Pleistocene Human Emergence*, edited by Erik Trinkaus. Cambridge University Press, in press.

© 1989 SCIENTIFIC AMERICAN, INC

Synthetic Zeolites

In the past 30 years man-made analogues of porous minerals called zeolites have revolutionized aspects of the petroleum industry. The search for novel commercially applicable forms continues apace

by George T. Kerr

In 1756 Baron Axel F. Cronstedt, a Swedish mineralogist, made the astonishing observation that certain minerals, when they were heated sufficiently, bubbled as if they were boiling. He called the substances zeolites, from the Greek words *zeo,* to boil, and *lithos,* stone. Since then approximately 40 mineral, or natural, zeolites have been discovered.

Analyses of chemistry and structure have shown that the "boiling stones," which are now known to consist primarily of silicon, aluminum and oxygen and to host an assortment of other elements, are highly porous crystals, veined with submicroscopic channels. The channels contain water (which accounts for the bubbling at high temperatures), but the water can be eliminated—by heating combined with other treatments—without altering a crystal's structure. Water-free zeolites can serve many purposes, including acting as molecular sieves: "reverse filters" that, in contrast to typical filters, capture relatively small molecules and allow larger molecules to flow by unimpeded.

By the early 20th century the unusual properties of zeolites had stimulated chemists to try to synthesize the crystals. Many succeeded, even inventing structures not found in nature. The synthetic forms have now become so important commercially that they are produced in greater quantity than most other crystalline materials.

GEORGE T. KERR, a member of the international editorial board of the journal *Zeolites,* investigated the properties and synthesis of zeolites at the Mobil Research and Development Corporation from 1956 until his retirement in 1985. He earned his Ph.D. in organic chemistry from Pennsylvania State University in 1952, whereupon he began a career in industry. When Kerr joined Mobil, he became one of the company's first employees to devote full-time efforts to the study of zeolites.

Indeed, certain synthetic zeolites have revolutionized aspects of the petroleum industry. Most notably, the worldwide production of gasoline by the catalytic cracking, or breaking, of large hydrocarbon molecules in petroleum now depends on the catalytic activity of zeolites. Crude petroleum is initially separated by distillation into fractions ranging from light products such as methane, propane and gasoline (which consist of small molecules) to heavier products such as gas oil and tar (which consist of heavier molecules). The molecules in gas oil are then routinely cracked to form what is called synthetic gasoline.

Zeolite catalysts yield from 10 to 20 percent more gasoline than earlier catalysts did and do so more cleanly and at lower, energy-saving temperatures. It is no wonder, then, that investigators continue to seek new forms that will have novel commercially exploitable features.

Any given zeolite is distinguished from other zeolites by structural differences in its unit cell: the smallest group of atoms that together exhibit all the chemical and physical properties of a crystal. The framework of every zeolite unit cell is constructed from tetrahedral building blocks, which can arrange themselves in varied combinations.

Oxygen atoms establish the four vertexes of each tetrahedron. These atoms are tightly bound to, and together enclose, either a silicon (Si) or an aluminum (Al) atom. The vertex oxygens are each shared by two tetrahedrons, so that every silicon atom or aluminum atom within the tetrahedral cage can be said to be bound to its four neighboring caged atoms through an intervening oxygen. A law of nature known as Loewenstein's rule prevents two aluminum atoms from sharing the same oxygen, and so the number of aluminum atoms in a unit cell is always smaller than, or at most equal to, the number of silicon atoms.

Before the aluminum atoms join the framework, they are actually ions that lack three of their original electrons. As a result, they can readily accommodate electrons donated from three of their bound oxygen atoms, but the electron donated by the fourth oxygen imparts a negative, or anionic, charge to the aluminum atom. This negative charge is balanced by a cation, or positively charged ion, from the alkali metal or the alkaline earth groups of the periodic table—commonly sodium (Na^+) or potassium (K^+) in the first case, calcium (Ca^{2+}) or magnesium (Mg^{2+}) in the second. The cations lie close to the anions but are in the crystal's channels, whose diameters range, in the natural minerals, approximately from 2.5 angstrom units (10-millionths of a millimeter) in sodalite to eight angstrom units in faujasite.

The cations play a major role in many zeolite functions and help to attract water molecules, which have no overall electrical charge but are polar: one end has a slightly negative charge, and the other end has a slightly positive charge. The cations are not a part of the framework, however, and under the right conditions they can be exchanged for other cations. Such an exchange has no effect on the crystal's structure.

The ion-exchange facility of mineral zeolites was perhaps the earliest commercially significant property to be recognized. Its discovery in the mid-1850's eventually led to the exploitation of the minerals as water softeners. Calcium and magnesium ions cause water to be hard, that is, unable to form a lather when mixed with soap. When such water is passed through zeolites that incorporate sodium cations, the zeolites give up the sodium and take in one calcium or magnesium ion for every two sodium atoms surrendered. (Because the aluminum in zeolites is prone to attack

© 1989 SCIENTIFIC AMERICAN, INC

by acidic tap water, zeolite water softeners have in general been replaced by more acid-resistant materials.) In another application the zeolite mineral clinoptilolite has been enlisted as an ion exchanger to extract radioactive elements from nuclear waste.

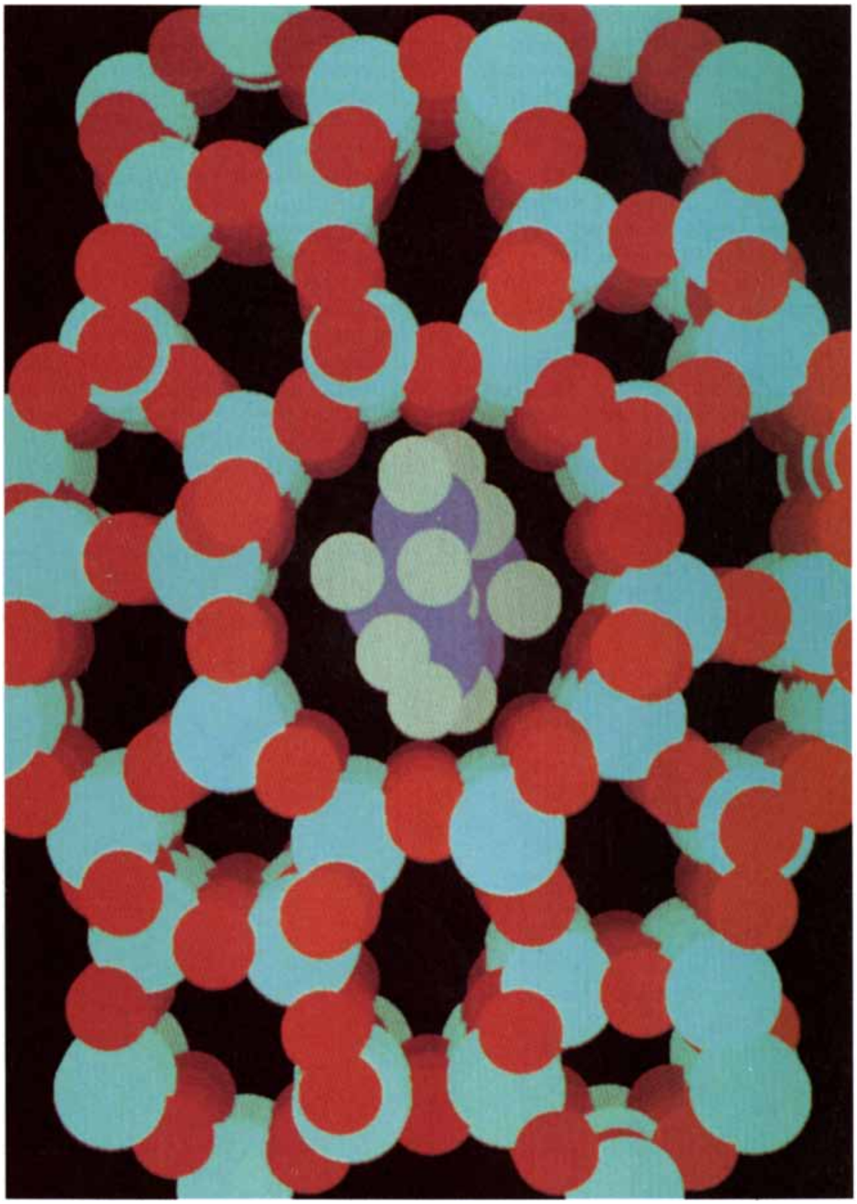
It was in 1932 that James W. McBain of Stanford University focused attention on another asset of zeolites: their filterlike character. The zeolite chabazite had been observed to adsorb—to take up and retain—the vapors of such relatively small molecules as those of water, formic acid and methyl and ethyl alcohols but to adsorb essentially no acetone, ether or benzene, whose molecules are larger. McBain recognized that the selectivity was influenced by the size of the crystal's pores, and it was he who coined the term molecular sieve. Chabazite's molecular-sieve properties have been exploited to prevent pollutants such as sulfur dioxide gas from escaping into the atmosphere through smokestacks.

In the years since 1932 natural zeolites have found many other applications as well. For instance, they can be excellent desiccants, or drying agents: when their water is removed, they readily adsorb water or other polar fluids from their environment. Still, in the past half century natural zeolites have been exploited less than the synthetic varieties, primarily because for many years they were thought to be rare. Now it seems that some mineral zeolites are actually abundant, and interest in them is growing again.

The most prodigious and systematic early studies of synthetic zeolites were led by Richard M. Barrer, who did much of his work at the Imperial College in London. His efforts in the late 1930's and beyond resulted in the preparation of a number of zeolite structures resembling those of natural minerals and the synthesis of a few structures not known in nature.

Barrer's work did not directly lead to commercially valuable products, but it did stimulate investigators in the Linde Division of the Union Carbide Corporation to launch in 1948 what turned out to be groundbreaking studies of zeolite synthesis. The company was already in the business of separating the components of air by distillation, and Robert M. Milton and the late Donald W. Breck wondered if molecular sieves would do the task more efficiently [see "Molecular Sieves," by D. W. Breck and J. V. Smith; SCIENTIFIC AMERICAN, January, 1959].

They departed from Barrer's experimental approach. He had worked with reaction mixtures and conditions simulating those that give rise to mineral zeolites, which form only in the presence of water and at temperatures of 100 degrees Celsius or higher. Deposits are most often found near volcanoes, sites that provide heat as well as silicon, aluminum and oxygen in the form of alumina (Al_2O_3) and silica (SiO_2). These minerals dissolve in the extremely basic, or caustic, solution (pH greater than 9) that forms when rainwater washes alkali metal and alkaline earth elements into groundwater from rocks; over time the elements aggregate to form crystals. Consequently, Barrer had exposed similar



MODEL of the zeolite crystal ZSM-5, shown with a molecule of the hydrocarbon para-xylene (*gray and blue shape*) in one of the crystal's submicroscopic channels, highlights the porosity found in all zeolites. ZSM-5 is exploited for the "shape-selective catalysis" that converts toluene (a by-product of petroleum refining) into para-xylene and benzene, both of which are commercially important. Toluene molecules that enter the crystal are altered to yield several products, but only benzene and para-xylene are small and compact enough to diffuse through the channels and escape; the other products remain trapped until they in turn are appropriately altered. The red spheres represent oxygen atoms, the light-blue spheres aluminum or silicon atoms. Cations (positively charged ions), which are responsible for catalysis, are also present but are too small to depict. The image was produced with Chem-X software, developed and distributed by Chemical Design Ltd., in Oxford, England.

© 1989 SCIENTIFIC AMERICAN, INC

mixtures to elevated temperatures (as high as 400 degrees C) and pressures, both of which he thought were necessary for zeolite formation. The resulting crystals had channels too small to separate air, however, which led Milton and Breck's group to see whether they might attain better results with slightly different reaction mixtures and with lower temperatures and pressures.

Then, as now, workers could play with the ratios of materials in their starting mixtures and with temperature and pressure to create zeolites of different composition or structure (or both)—and they could replicate their recipes—but they could not actually predict what would be achieved by changing a recipe in a given way. Yet in one instance Milton and Breck made a zeolite that had a novel structure and later proved to be quite valuable. They called it Linde A.

Linde A, which contained only sodium cations, was not initially effective for separating air components, but when the large, channel-clogging sodium ions were exchanged two for one with calcium cations, the altered crystal did become quite adept at separating nitrogen and oxygen. It was also found that the calcium form of Linde A could remove certain undesirable hydrocarbon molecules from gasoline.

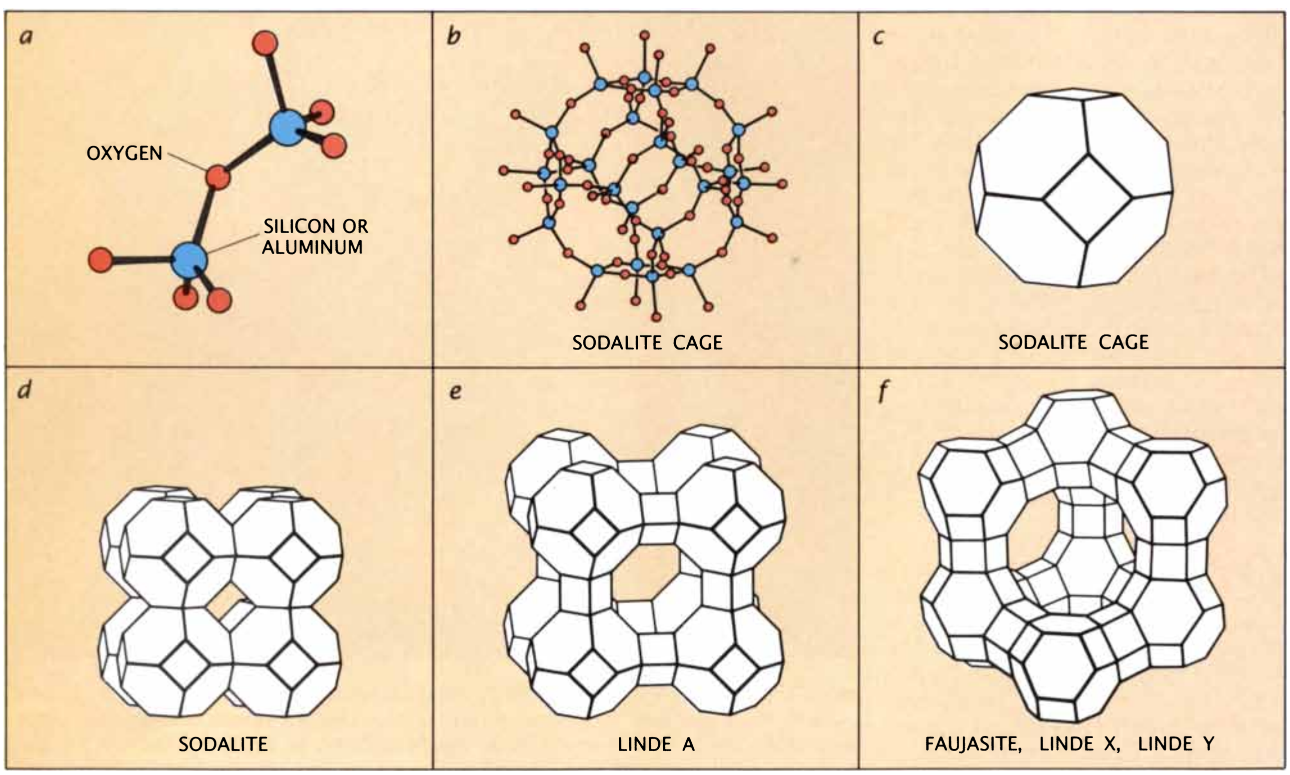
When natural gasoline is separated from crude oil by distillation, it contains various hydrocarbons that are similar in size but different in shape. Some consist of straight molecular chains, whereas others are cyclic (ring-shaped) or branched. The shapes of the constituents matter because the straight-chain fraction burns more explosively—it "knocks"—and thus interferes with smooth engine performance. The calcium form of Linde A managed to adsorb the straight-chain hydrocarbons, leaving behind a higher-quality product.

The new zeolite was first exploited commercially in the mid-1950's and continues to be employed to separate air components and to improve gasoline today. Linde A has also been employed as a desiccant—for example, to remove water from the hydraulic brake systems of trucks and buses.

At about the same time as the Union Carbide workers developed Linde A, they also produced Linde X, which was to become the first zeolite to perform commercially as a catalyst. This material had the same structure as the rare large-pore mineral faujasite but had a different ratio of silicon to aluminum and, like Linde A, could be synthesized directly only in the sodium-cation form.

At first Linde X seemed to have no obvious applications, but when Union Carbide made samples available, many investigators took advantage of the opportunity to study it. Among them was Robert W. Schiessler of what is now the Mobil Research and Development Corporation. He set in motion not only studies of Linde X but also in-depth studies of zeolites in general. Schiessler hoped workers at Mobil would synthesize zeolites having novel compositions and structures and, more specifically, would discover ways to make zeolites helpful in forming synthetic gasoline.

That objective seemed feasible be-



FRAMEWORK of zeolites is constructed of tetrahedral building blocks; in some cases there is a characteristic "sodalite cage." The four tetrahedral vertexes are filled by oxygen atoms (*orange*), and a silicon or aluminum atom (*blue*) lies at the center (*a*); each oxygen is shared by two tetrahedrons. In representations of the sodalite cage (*b*), which consists of 24 tetrahedrons, and of zeolites in general, the ball-shaped atoms are usually omitted (*c*): straight lines join the centers of two tetrahedrons, so that each vertex represents an aluminum or silicon atom and the midpoint of each line represents an oxygen atom. Sodalite cages are found in all of the structures depicted here: sodalite itself (*d*), the synthetic crystal Linde A (*e*) and a group of like-structured crystals, namely the mineral faujasite and the synthetic zeolites Linde X and Linde Y (*f*).

© 1989 SCIENTIFIC AMERICAN, INC

cause the catalyst then responsible for the cracking of gas oil was chemically similar to zeolites: it included silicon, aluminum, oxygen and, at least in its initial postsynthesis state, sodium. Unlike zeolites, however, the catalyst was amorphous.

I had joined Mobil in 1956, and soon after that Edward J. Rosinski of the company and I set out to try to activate Linde X (whose channels were conceivably large enough to allow hydrocarbons to enter) in much the same way as workers normally activated the amorphous material. They bathed it in an aluminum-rich solution and thus exchanged the sodium cations, which are "poisons" that interfere with catalysis, for aluminum cations (Al^{3+}). This exchange essentially converted the amorphous material into an acid—something that supplies hydrogen ions (H^+), or single protons. Such acidity is crucial to the catalytic cracking of hydrocarbons, although exactly how protons interact with the hydrocarbon molecules to crack them is still not clear. (Apparently the hydrogen ions derive from water molecules in the gas oil and form when the aluminum cations detach hydroxyl (OH^-) groups from the water.)

The aluminum bath seemed like a good idea for the zeolite, but the treatment merely plugged its channels. Later our colleagues Jean M. Bourguet and F. David Hart also decided to tackle the activation problem. At about that time I recalled that ions of the rare earth element cerium (Ce), which like aluminum cations have a charge of 3+, could be exchanged for the sodium ions in Linde X, and so I suggested they try that exchange.

Hart did what he thought was the suggested experiment, although it turned out that the substance he had purchased as a source of pure cerium was mislabeled and included a mixture of rare earth elements instead. The error was immaterial, however, because the experiment succeeded in yielding a highly active catalyst anyway. In fact, to everyone's chagrin, it cracked the hydrocarbons too severely, producing methane and other light hydrocarbons rather than gasoline.

Rosinski and Charles J. Plank of Mobil came to the rescue by suggesting that steaming of rare earth zeolites might reduce their catalytic activity to the desired level. (Steaming can also increase such activity, but intensive steaming is destructive.) By 1962 our work resulted in the commercial release of the first zeolite-based gas oil cracking catalyst, Durabead 5, which yielded much more gasoline than the standard amorphous catalyst.

Exactly why Durabead 5 proved so much more effective is not fully understood even today, but there are several possible explanations. For instance, it seems reasonable to guess that the large, uniform channel system in the zeolite provided more surface area on which the molecules in gas oil could interact with protons.

Durabead 5 was a milestone in the history of synthetic zeolites, but its life as a commercial product lasted only until 1964. In the late 1950's Breck had synthesized another faujasite-type structure called Linde Y, with a silicon-to-aluminum ratio that ranged from 1.5 to 3.0—higher than the ratio in Linde X. When workers at Mobil devised a way to exchange rare earth cations for the sodium cations of the initially synthesized crystal, the rare earth form of Linde Y quickly replaced Linde X. Its higher silicon content and correspondingly lower aluminum content made it more durable. Framework aluminum atoms are the weak link in zeolite catalysts because at high temperatures they are less stable than silicon atoms and are prone to attack by water.

In the 1970's the rare earth form of Linde Y too was replaced, this time by another form of Linde Y. The newer variety remains the preferred cracking catalyst for gas oil today, because it is significantly more stable and less susceptible to destruction by water when it is exposed to high temperatures.

The greater resilience of this so-called ultrastable zeolite stems from changes in the crystal framework that occur during postsynthesis processing. Workers initially exchange the sodium cations in typical Linde Y crystals with ammonium ions (NH_4^+) and then heat the crystals. The heating causes ammonia gas (NH_3) to form and escape from the crystals, leaving behind hydrogen ions. As I discovered in the mid-1960's, steaming of this hydrogen form of the zeolite drives a fraction of the hydrogen ions from the crystals and causes an equal number of aluminum atoms in the skeleton to pop out of the framework, whereupon they each bind with eight oxygen atoms, forming what are called octahedral cations.

The benefits of the new configuration are twofold. The aluminum atoms are somehow replaced in the framework by more durable silicon atoms. Also, the aluminum atoms themselves become more resistant to heat and water in the octahedral conformation. The steamed zeolite is able to retain its catalytic power even though a number of hydrogen ions have escaped, because the octahedral aluminum cations play much the same catalytic role as they do in the amorphous catalyst (and as rare earth ions play in Linde X and the previous version of Linde Y).

Late in the 1950's several of us at Mobil began experiments that would lead in the mid-1960's to the commercial release of still another synthetic zeolite, called ZSM-5. (The last two initials stand for Socony Mobil, part of the company's name at the time.) This zeolite has several applications and has recently become important in the conversion of methanol (wood alcohol) to gasoline.

The creation of ZSM-5 was the culmination of a series of discoveries that began when I thought of substituting so-called quaternary ammonium cations for all or some of the conventional alkali metal ions in the reaction mixtures that give rise to synthetic zeolites. These are normal ammonium ions in which all four of the hydrogen atoms are replaced by organic groups. Their typical sources—quaternary ammonium hydroxides, or the ions bound by hydroxyl groups—are as strongly basic in solution as are sodium and potassium ions, but the quaternary ammonium cations are appreciably larger than either sodium or potassium cations. I therefore anticipated that the larger ions might induce the formation of zeolites whose structures were quite different from any others.

One ion I chose (tetramethylammonium, or TMA) resulted in a zeolite that George T. Kokotailo of Mobil then showed had the same crystal structure as Linde A, except that the unit cell was smaller and the silicon-to-aluminum ratio was a bit higher. We named it ZK-4, with the "K" referring to both of us. (At about the same time, Barrer and his student Patrick J. Denny described a similar structure prepared with the same cations.) Encouraged that the work was moving in the right direction, we then tested a more complex ion and succeeded in producing the first truly new zeolite structure synthesized with quaternary ammonium ions: ZK-5.

In collaboration with Rosinski and our colleague Robert L. Wadlinger, I then tried yet another ion: tetraethylammonium, or TEA. The result, zeolite beta, astounded us: it had a silicon-to-aluminum ratio that ranged from about 15 to 50, far surpassing the 5.5 ratio of mordenite, the most silicon-rich zeolite mineral. It was also some-

© 1989 SCIENTIFIC AMERICAN, INC

what hydrophobic (water-repellent), in stark contrast to natural zeolites and to the earlier synthetic varieties, all of which attract water strongly.

Many years later, in 1976, N. Y. Chen of Mobil related the two findings. He showed that as the silicon-to-aluminum ratio of zeolites increases (with a concomitant decrease in the number of aluminum ions and therefore of the cations available to attract water) the crystals become increasingly hydrophobic. Only within the past year has beta's structure been deciphered; not surprisingly, it has no known counterpart in nature.

In 1963 George R. Landolt, who worked with me for a few years, and Robert J. Argauer (also of Mobil) decided to see if they could further increase the silicon-to-aluminum ratio to produce a very silicon-rich zeolite. They could indeed, by substituting in the reaction mixture a quaternary ammonium ion that is three times as large as TMA. Thus, the zeolite ZSM-5 was invented. The selected ammonium ion, tetra-*n*-propylammonium (TPA), worked precisely because it was big. Only a limited number of large cations can fit into a zeolite framework; because of the need for charges to be balanced, this limitation severely constrains the number of associated aluminum ions that can be present, and the result is a silicon-rich structure.

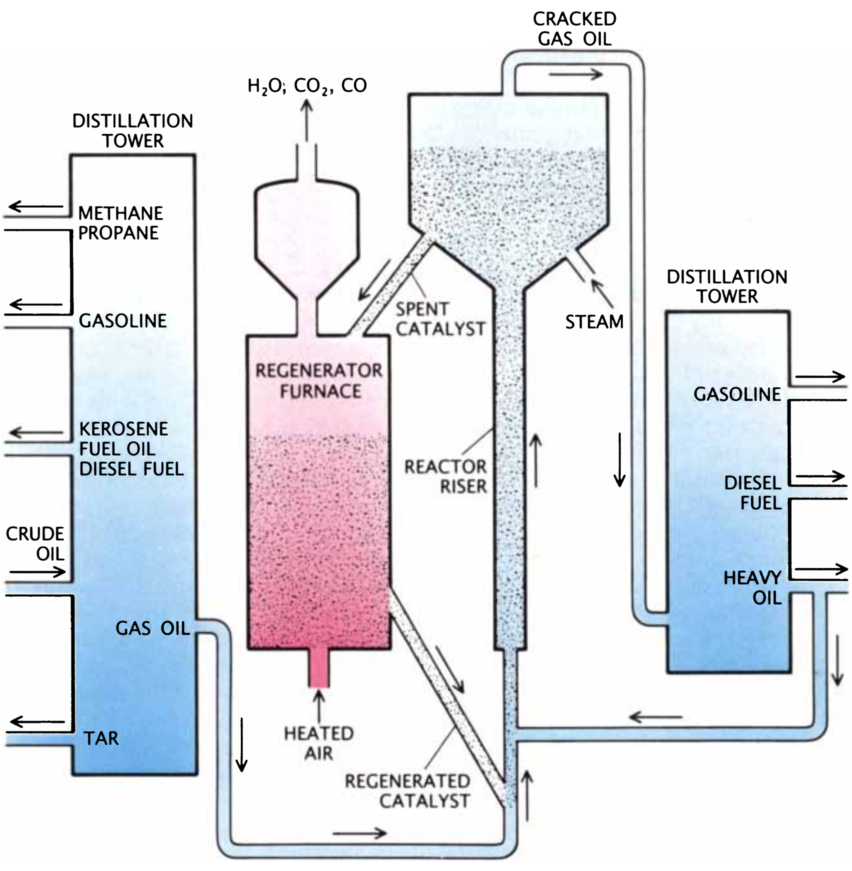
Several years afterward Francis G. Dwyer and Edwin E. Jenkins of Mobil managed to go still further, preparing varieties of ZSM-5 that had vanishingly small amounts of aluminum. Indeed, the framework of certain samples proved on careful chemical analysis to be more than 99.99 percent pure silica. Some of these extremely hydrophobic crystals have been enlisted to remove hydrocarbon contaminants from water.

On first inspection, the fact that ZSM-5 zeolites have few aluminum atoms and few cations might seem to preclude their acting as effective catalysts. Indeed, David H. Olson, Rudolph M. Lago and Werner O. Haag of Mobil have shown that in the hydrogen form of ZSM-5 (which is analogous to the hydrogen form of Linde Y discussed above) catalytic activity decreases in direct proportion to decreases in the amount of aluminum present (by weight) in the crystal. Presumably the catalytic activity and aluminum content are related because the number of aluminum atoms controls the number of protonic, or hydrogen-ion, sites.

Yet Olson, Lago, Haag, I and other Mobil workers—Richard J. Mikovsky, Stuart D. Hellring, Kirk D. Schmitt—have shown that, in fact, the ZSM-5 zeolites can be transformed into superactive catalysts. In particular, we have found that by exposing the hydrogen form of ZSM-5 to mild steaming we can generate new catalytic sites that are from 45 to 75 times more active than normal protonic sites.

Such steam-treated zeolites are not exploited for the catalytic cracking of gas oil, because their elliptical channels are too small to accommodate the large hydrocarbons in gas oil; the chords of their minor and major axes measure only 5.4 and 5.6 angstrom units. The size is perfect, however, for the catalysis of certain smaller, cyclic hydrocarbons. For example, toluene, a by-product of conventional petroleum refining, can be upgraded by ZSM-5, which transforms it into two related, but more valuable hydrocarbons, benzene and para-xylene. Benzene is a ubiquitous solvent that is found in paints, varnishes and cleaning agents; para-xylene is converted into terephthalic acid, which in turn is exploited in the production of a variety of synthetic materials.

The synthesis of large quantities of pure para-xylene is accomplished by an increasingly important zeolite trick that was first postulated by Vincent J. Frilette and Paul B. Weisz of Mobil in 1961: shape-selective catalysis. In this process the products yielded by catalysis are controlled in part by the size or shape (or both) of the zeolite's channels. Toluene molecules, which are small enough to penetrate the ZSM-5's channels, are passed through a bed of the zeolites and simultaneously heat-



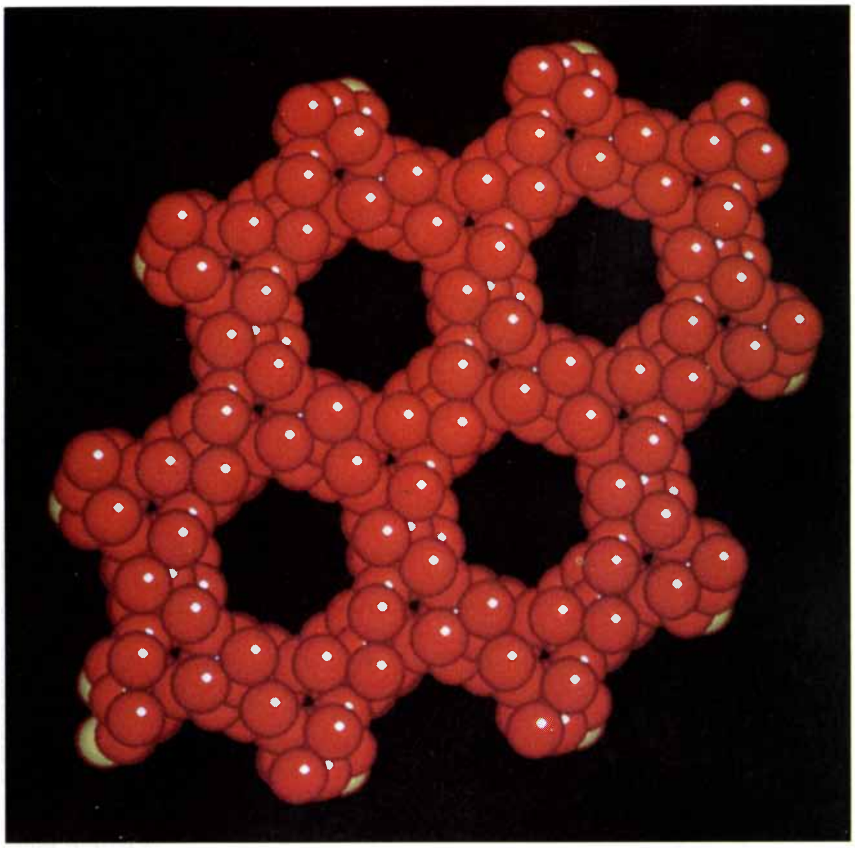
CATALYTIC CRACKING, or breaking, of hydrocarbons by zeolite crystals to produce gasoline and other products is done after crude petroleum is separated by distillation into a range of products from methane (which has small molecules) to tar (which has large ones). Some of the products are listed at the left. The gas-oil fraction is vaporized by contact with hot particles (*dots*) containing the zeolite Linde Y. Together with the particles, the vapor moves up a tube called a riser, where the hydrocarbons (which diffuse into the zeolite's pores) are cracked. Above the riser the vapor is passed into a distillation tower that separates the resulting fragments by size: smaller molecules form gasoline, and heavier ones form diesel fuel and what is called heavy oil, some of which is recycled. Meanwhile, the catalytic particles themselves are prepared for recycling. They are steam-treated to remove hydrocarbons and passed to a regenerator, where hot air burns off a carbonaceous deposit called coke.

© 1989 SCIENTIFIC AMERICAN, INC

ed to approximately 520 degrees C. At this temperature the zeolites catalyze changes in the toluene, yielding benzene and three different forms of xylene (dimethylbenzene).

Benzene and para-xylene both escape from the crystals readily, but the other two xylenes (ortho- and meta-), which are shaped differently than their cousin, have difficulty making their way out through the channels. While they are "stuck," they are likely to be converted into the para form, whereupon they escape. Thus, ZSM-5 selectively ensures that most of the xylene molecules diffusing out of it will be of the para variety.

How does ZSM-5 catalyze the shape-selective conversion of methanol to gasoline? This process is most significant for countries, such as New Zealand, that have no oil deposits but do have an abundant supply of methane (CH_4), or natural gas, which can readily be converted to methanol (CH_3OH). After the methanol is partially broken down into a compound called dimethyl ether and water, the ether is passed through the zeolite catalyst, where it is further converted to water and a complex mixture of hydrocarbons. As is true in the production of para-xylene, molecules larger than the desired gasoline hydrocarbons become trapped in the crystal's channels, diffusing out only when the catalytic process yields molecules that are appropriately small.



VPI-5, a newly synthesized crystal, is similar in structure to true zeolites, but its framework has little silicon; instead it consists primarily of phosphorus, aluminum and oxygen. Only the oxygen atoms (*red spheres*) are clearly visible in this model. VPI-5's channels are some 12 to 13 angstrom units (10-millionths of a millimeter) in diameter, bigger than those of any known natural or synthetic zeolite. At that size the channels are able to accommodate rather large hydrocarbons, which suggests that catalytic forms of VPI-5 may one day prove valuable for the cracking of petroleum fractions heavier than gas oil. Mark E. Davis of the Virginia Polytechnic Institute and State University, whose research group invented VPI-5, made the image.

Some of the most interesting zeolites developed since ZSM-5 became available are not zeolites at all, at least not in the strict sense of the term, which generally requires that the crystalline framework consist exclusively of silicon, aluminum and oxygen. The new materials, all of which are still laboratory curiosities, are based on phosphorus rather than on silicon. Some workers consider them to be zeolites, however, because they have uniform intracrystalline channels that are very similar to, and in some cases identical with, those of some zeolites. Several of the new materials also have classical zeolite structures and adsorptive properties.

The promise of these materials—as well as ones that are altered to resemble more closely the chemistry of true zeolites—arises from the fact that some of them have novel structures that may enable them to be applied as catalysts, molecular sieves, adsorbents or ion exchangers in situations where no zeolite has yet proved especially helpful.

For instance, Mark E. Davis and his colleagues at the Virginia Polytechnic Institute and State University have synthesized a zeolitelike aluminum phosphate crystal, VPI-5, whose channels are larger than any found in natural or synthetic zeolites: some 12 to 13 angstrom units in diameter. The large size raises the possibility that some version of the crystal may one day be exploited as a catalyst to crack larger hydrocarbons than Linde Y can handle, thereby enabling manufacturers to extract more gasoline from their petroleum sources.

What does the future hold? In addition to the surprises still to be revealed by experiments with aluminum phosphate "zeolites" and their derivatives, I suspect that important new structures may emerge from other experimental strategies. Recently, for example, investigators have prepared zeolites from organic, nonaqueous solvents instead of from the usual inorganic, aqueous solutions.

What is still needed, though, is insight that will enable chemists to predict the structures and other properties that will emerge when various starting ingredients are mixed together in specific ratios at particular temperatures and pressures. Until then trial and error will prevail. Fortunately, that approach has served industry rather well so far.

FURTHER READING

ZEOLITE MOLECULAR SIEVES: STRUCTURE, CHEMISTRY AND USE. Donald W. Breck. John Wiley & Sons, Inc., 1974.

ZEOLITE CHEMISTRY AND CATALYSIS. Jule A. Rabo. ACS Monograph 171, American Chemical Society, 1976.

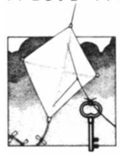
HYDROTHERMAL CHEMISTRY OF ZEOLITES. R. M. Barrer. Academic Press, 1982.

THE ZEOLITE CAGE STRUCTURE. J. M. Newsam in *Science*, Vol. 231, No. 4742, pages 1093-1099; March 7, 1986.

© 1989 SCIENTIFIC AMERICAN, INC

THE AMATEUR SCIENTIST

What do phonograph records have in common with windshield wipers?



by Jearl Walker

Position an ordinary phonograph record and a small desk lamp on a table so that the light from the lamp shines on the record from the side opposite you in an otherwise dark room. With the center of the record about midway between you and the lamp, view the record with one eye closed and from the height of the bulb in the lamp. The entire record is illuminated, but does its entire surface appear bright? Or is there a single bright spot like the one you would see in a mirror?

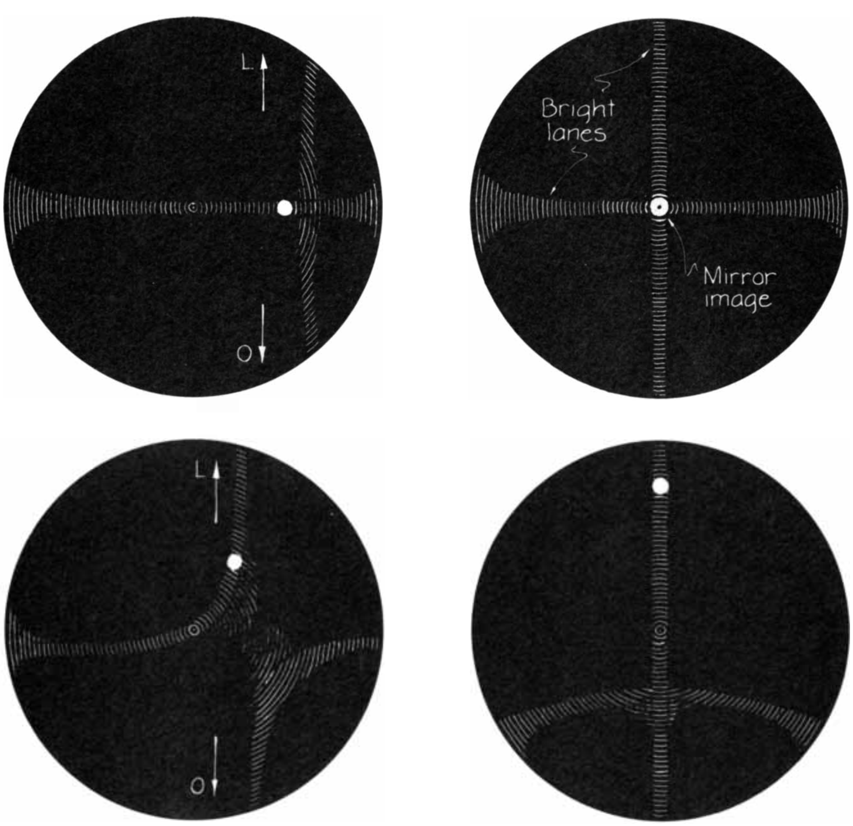
Neither of the above. Instead there are bright, narrow lanes in the form of a cross. The rest of the record is dim (it would be dark but for the light that is scattered to it from the walls and ceiling and objects in the room). One arm of the cross runs along the line between you and the lamp, and the other arm passes through the first at a right angle [*see drawing at top right in illustration below*]. The figure assumes that the center of the record lacks the usual label and central hole and that it is covered with grooves just as the rest of the record is. If that were actually the case, you would see an especially bright spot at the center of the record.

If you shift the record to the left, the bright lane between you and the lamp begins to curve slightly, but the other lane remains straight and still extends horizontally through the center of the record, as is seen in the top left drawing. Provided that the record is not shifted too much, the especially bright spot lies on the straight lane, near the other lane. One end of the curved lane points toward the lamp (that end is labeled *L* in the illustration), and the other end (labeled *O* for "observer") points toward you. If instead you shift the record toward you from its original central location, it is the left-to-right lane that curves gently; the other lane remains straight, as in the bottom right drawing. The bright spot now lies on the straight lane on the opposite side of the record from you.

When the record is moved both toward the left and toward you, the lanes separate into shapes that resemble hyperbolas, as in the bottom left drawing. One lane still passes through the center of the record; if the record is not shifted too much, the bright spot is still on that lane. The second lane lies near an edge of the record; it enters at one point on the edge, heads for the center and then swerves toward another point on the edge. The pointing is now divided between the lanes: one end of the first lane points toward the lamp, and one end of the second lane points toward you. If the record is shifted farther, the second lane shrinks toward the edge and then finally disappears.

The width of the lanes depends on the size of the bulb and the nearness of the lamp. A large bulb in a lamp close by creates wide, rather indistinct lanes. You can improve the visibility of the patterns if you cover the opening of the lamp with a sheet of aluminum foil through which a small hole has been punched. Then the lanes are narrow and clear. (Take care that the lamp does not overheat.)

What accounts for these patterns? The answer was provided in 1963 in an impressive study by J. B. Lott, who had just graduated from the Felsted School in Essex, England. It was clear, of course, that the patterns are reflections from the grooves on the record; the problem was to determine which parts of which grooves have the proper orientation to reflect light to an observer. Lott's solution was to imagine that the lamp and the observer are inside an ellipsoid: a dirigiblelike



Reflection patterns from a phonograph record

© 1989 SCIENTIFIC AMERICAN, INC

structure generated when an ellipse is rotated about one of its axes. Wherever part of a groove happens to coincide with the side of an ellipsoid, light reflects to the observer. I have found that the analysis also explains a more common observation: the bright streak produced by a streetlamp seen at night from a car through a windshield that is being swept clear of rain or snow by the wipers.

To follow Lott's analysis, you first must understand how light reflects from a flat surface. Suppose that the record lacks grooves. Also suppose that the bulb in the lamp is small enough to be considered a point. Light rays from the lamp spread over the record and reflect from the surface. A line that is perpendicular to the surface at a point of reflection is called a normal, and the angle of the ray from the lamp is measured relative to it [*see top left illustration on this page*]. The ray is called the incident ray, and its angle is the angle of incidence. The ray that leaves the surface is called the reflected ray, and its angle with the normal is the angle of reflection.

The reflection obeys two simple rules. The normal, the incident ray and the reflected ray must lie in a plane, and the angle of reflection must equal the angle of incidence. If the center of a smooth record lies halfway between you and the lamp, you intercept a ray that reflects from the center. All the other rays reflect from the record surface but not in the direction of your eye. What you perceive is a single spot of light at the center of the record, as if the spot, rather than the lamp, were the source of the light.

I shall call the spot the "mirror image" of the lamp, because in producing it the record functions as a mirror. When your eye is at the height of the lamp, the mirror image is always halfway between you and the lamp. You can control its location on the record by shifting the record over the table or (equivalently) by moving your position. For example, if you shift the record to the left and also toward you, the mirror image appears in the upper right quadrant of the record. (The mirror image is, of course, the especially bright spot that, as I mentioned above, is seen in the pattern of reflections from a grooved record.)

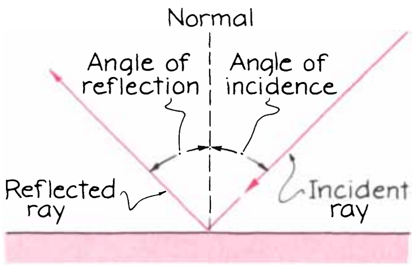
The rules of reflection also apply to a surface that is smoothly curved, but in this case the normal is perpendicular to a line that is tangent to the surface at the point of reflection, as is shown in the top right illustration on this page. Lott assumed that the grooves on a record are a series of smoothly varying hills and valleys that form circles around the center of the record. If a radial cross section is taken through any one of the hills, and if normals are erected at points along the hillsides in the section, the hill bristles with normals—somewhat like the quills on an angry porcupine. The normal at the top of the hill is vertical. All the others are tilted radially from the vertical, either toward the center of the record or away from it.

How light reflects to an observer from the grooves can be described in terms of the normals. If a ray is to reach the height of the observer's eye, the normal at the point of reflection must have a particular tilt. Lott argued that this requirement is easily met by every groove. Pick a spot around the circle of a groove, and then consider the normals along the sides of a cross section through the hill. At least one of the normals will have the proper tilt for the light reflecting at the foot of the normal to be sent to the height of the eye. The requirement that a ray be sent to a certain height, then, is not responsible for the patterns seen on a record.

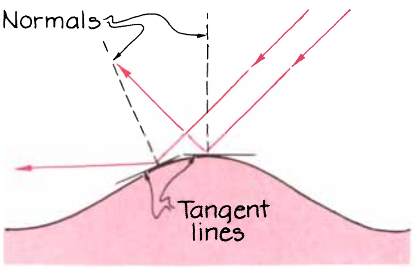
Lott reasoned that another, more restrictive requirement is responsible: the reflected ray must be directed properly, left or right, toward the eye. Only certain spots on certain grooves meet that requirement. The argument greatly simplifies the analysis, because it reveals that only the horizontal orientation of the normals to the circles formed by the grooves is important. The analysis is further simplified by the fact that horizontally the normals are radial: they can be extended through the center of the record. In what follows I shall ignore the vertical aspect of the normals and concentrate on their radial aspect.

To find which segments of which circles produce the patterns, Lott relied on a special property of an ellipsoid. If a light ray originates at one of the two focal points in an ellipsoid and reflects from the interior, it must pass through the other focal point. Lott imagined that when an observer receives a reflection from a record illuminated by a lamp, the lamp is effectively at one focal point of an ellipsoid and the observer at the other focal point. The record cuts horizontally through the lower part of the ellipsoid, and in the slice that is cut out, the wall of the ellipsoid forms an ellipse [*see illustration below*].

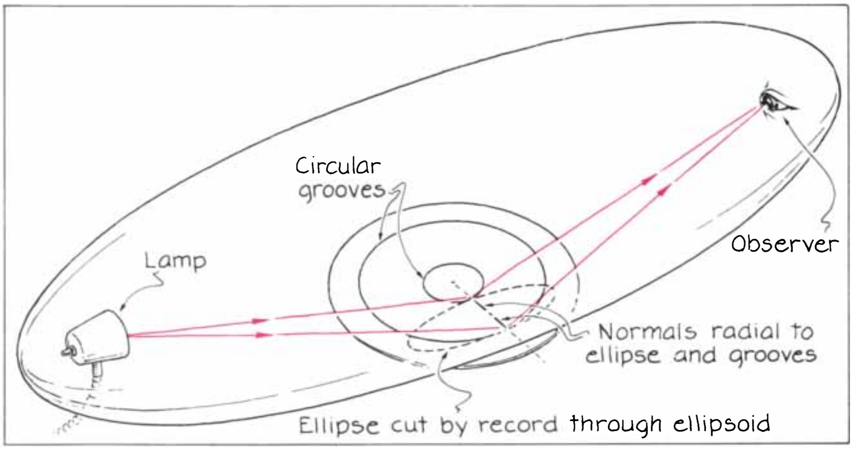
If the light is to reflect from the ellipsoid, it must reflect from the part of the wall that forms the ellipse. Because it must also reflect from the



A reflection from a flat surface

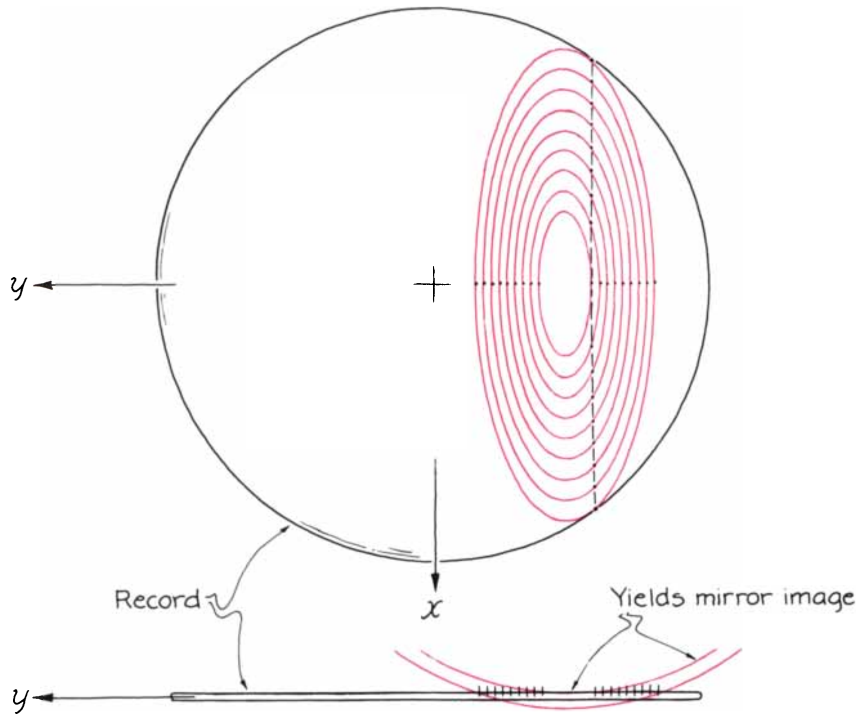


Reflections from a curved surface



The light is reflected from an elliptical section of an ellipsoid

© 1989 SCIENTIFIC AMERICAN, INC



Ellipses left by ellipsoids

has been shifted from its initial central location directly to the observer's left. The top figure in the illustration is an overhead view of the ellipses that are left by ellipsoids in the plane of the record. The *y* axis runs to the observer's left and right, and the *x* axis is perpendicular to the *y*. The bottom figure is a vertical cross section taken along the *y* axis. Included in it are the lowest parts of cross sections through two of the ellipsoids.

To construct the overhead sketch, you begin with the mirror image, which is the especially bright spot in the reflection patterns. Recall that when the observer's eye and the lamp are at the same height the mirror image is halfway between them. In the present case, that condition puts it on the *y* axis to the right of the record's center. The image is associated with an ellipsoid that just barely touches the top surface of the record, and so it leaves an ellipse that is as tiny as a point in the plane of the record. A segment of that ellipsoid is shown in the bottom figure in the illustration. Draw that point on an overhead sketch.

circle of a groove, the circle and the ellipse must coincide at the point of reflection. That in turn means that the circle and ellipse share a common normal there. Because the normal to any part of a circle is radial to the center of the record, the point on the ellipse that reflects light to an observer is a point with a radial normal.

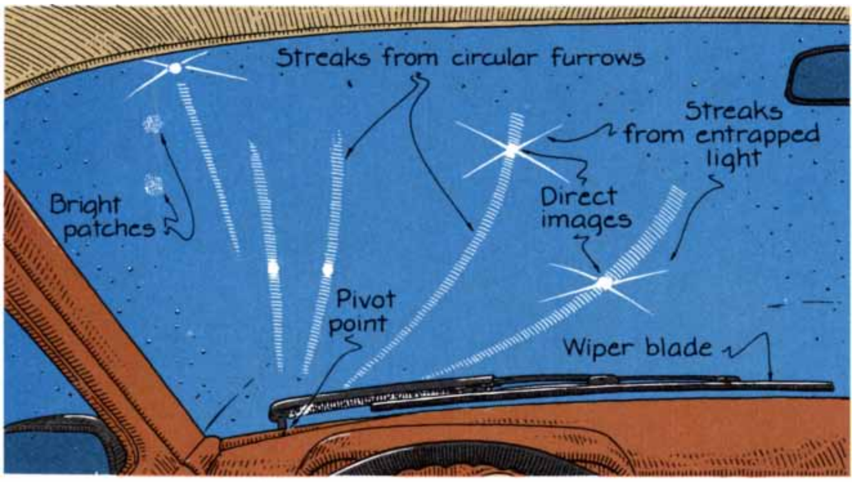
To account for a full pattern of reflections on a record, one needs to consider a family of ellipsoids of different sizes. The record's slice through some of the ellipsoids leaves a large ellipse, and its slice through others leaves a small ellipse. In each case the points on an ellipse that are responsible for the reflection patterns are points with radial normals. These conditions were sufficient for Lott to solve equations for ellipses, circles and radial normals in order to find the points on a record that reflect light to an observer's eye. He restricted the situation to the one in which the lamp and the eye are at the same height above a horizontal record. The composite of the reflection points forms the lanes seen on the record.

I found that the general shape of the lanes could be predicted with a series of sketches without the need of equations. An example is shown in the illustration above, in which the record

Around the point draw a small ellipse centered on the point and with its long axis parallel to the *x* axis. The ellipse is associated with an ellipsoid that is larger than the first one and extends through the record; this ellipsoid and its ellipse have the same orientation. Now consider the normals at points around the ellipse. (Keep in mind the fact that a normal is perpendicular to a tangent to the ellipse.) Mark those points that have a normal that is radial—that is, one that can be extended through the center of the record. There are four such points: two are on the *y* axis (one on each side of the mirror image), and two are near the sharply curved ends of the ellipse.

Now repeat the procedure with similar but progressively larger ellipses derived from progressively larger ellipsoids. Each ellipse has four points with radial normals. After you have located a number of the points, connect them with lines; the lines replicate the bright lanes in a reflection pattern. If you tilt the record or bend it gently, the reflections are more difficult to explain, but the shapes of the patterns are similar.

When I came across Lott's paper and first took notice of the reflection patterns on a record, I suddenly felt I had seen one of the patterns in a different setting. Then I realized that whenever I drive through rain or snow at night and look at a streetlamp or the headlamp of an oncoming car through a



Streaks and patches of light in a wet windshield

© 1989 SCIENTIFIC AMERICAN, INC

wiper-cleared area of the front windshield the lamp has a streak running through it [*see bottom illustration on opposite page*]. Might that streak be related to one of the bright lanes on a record?

The streak of light in a windshield was discussed in 1954 in a brief note by Paul Kirkpatrick of Stanford University. He attributed it to the fact that when the rubber blade of a windshield wiper rubs across the glass it scrapes furrows in gummy road debris that tends to adhere to the glass. Because the wiper moves in an arc, the furrows form circles around the pivot point of the wiper; the furrows may persist long after the windshield is dry.

When you look through the array of circles and toward a lamp, you intercept light that reflects from the sides of the furrows. In daytime the light is too dim to be visible, but at night it can be so apparent as to be annoying. The streak is often straight, with one end pointing toward the wiper's pivot. As your angle of view of the light source changes, the direct image of the lamp slides over the windshield, and the streak rotates around the image so that one end continues to point toward the pivot. Kirkpatrick mentioned that sometimes the streak is noticeably curved, with the lower end still pointing toward the pivot. Sensitized by Lott's analysis, I wondered whether the streak might be related to a bright lane on a record—the one on which the mirror image is seen.

After a little thought I realized that the situations were almost identical. The display I see in the windshield is essentially the one I see on the far side of a record when I move the record toward me and thus shift the mirror image to the far side, as in either of the bottom drawings in the illustration on page 106. In both cases a bright line extends through the center of the circles, or at least points there, and the line can be straight or curved depending on the circumstances. Somewhere along the line there is an especially bright spot. With the record, the bright spot is a mirror image and the light source is on my side of the circular array. With the windshield, the bright spot is a direct image and the light source is on the opposite side of the array. I concluded that Lott's analysis applies also to a windshield streak.

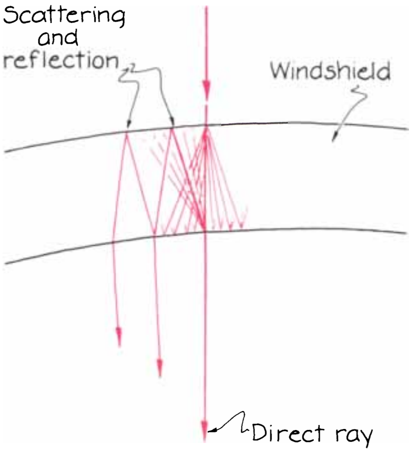
Still, one detail nagged me. With the windshield, the lack of "near quadrants" (corresponding to the near side of a record) apparently eliminated the second bright lane or streak; any second streak (which would run left to right or would be strongly curved and separate from the first one) would presumably be below the pivot point of the wiper. Was there some way I could bring the second lane into sight? I tried and failed several times to make the streak appear. Then one day, with the setting sun seen through the curved section of the windshield on the passenger's side, I saw both bright streaks near the sun; they resembled the bottom left drawing in the illustration on page 106. (If you watch for sun streaks, be careful about looking directly toward the sun for more than a second because of the obvious danger to your eyes.) I found that by altering my perspective I could make them join or separate at will—and in the process I almost drove off the road. I still do not know exactly why that second streak appeared.

During the past winter I enhanced the visibility of the streaks by allowing the salt that is spread to clear roads of snow and ice to build up on my windshield. The furrows ground out by the windshield wiper were then quite apparent. The streaks from them were probably due more to a complex scattering of the light than to simple reflections as envisioned by Kirkpatrick, but the results were similar. Once I noticed a novel feature of the streaks. As I drove in the direction of a low sun, the bare limbs of trees blocked part of the sun's image and radiated dark lines down through the sun streak.

When a car approaches you at night, the extension of the streaks from its headlamps toward the pivot point of your wiper forces the streaks to converge. Each streak also narrows as it nears the pivot, because the circles nearer the pivot are more strongly curved than the circles farther out.

Kirkpatrick mentioned an illusion of depth that can be perceived in the streak from a streetlamp. Because of the separation of your eyes, each eye sees its own streak. The separation between the streaks is greatest near the pivot and is least at the direct image of the lamp. You might perceive both streaks, but if your brain successfully merges them into a composite, their actual separation creates the illusion of depth: you perceive a single streak that appears to be a lit path extending from you to the lamp. If the streaks are gently curved, the path is curved too, as if it leads over an invisible valley and then up an invisible hill to the lamp. (Similar depth can sometimes be seen in a reflection pattern on a record.)

Most windshields are curved, particularly at the sides. The curvature alters the shape of the streaks just



The momentary entrapment of light

as the bending of a record changes a reflection pattern. The curvature also allows regions of gummy material to scatter light to you that would miss you if the glass were flat. It also momentarily traps some of the light from a lamp, as is shown in the illustration above, before the light is sent in your direction. During the entrapment the light reflects and scatters many times inside the glass, moving away from the direct image of the lamp. What you see is a bright streak that extends from the direct image. If you reach out the car window and block the direct image with a finger, the streak disappears. Similar streaks arise from corrective eyeglasses, but in a car they can be distinguished from the windshield streaks if you tilt your head: the streaks from the glasses rotate.

Here are a few more questions raised by reflections from circular arrays. If you shift a record enough in the light of a lamp, why does the bright lane through the center develop an end at which the reflection dims and then disappears? (A hint: recalling Lott's assumption that there is always a normal on a groove that sends a ray to the height of your eye, consider just how tilted a normal can be at the juncture of a hill and valley.) What sort of reflections are seen in a "laser disk"? Why are the reflections on both a record and a laser disk colorful when the lamplight is white?

FURTHER READING

A BINOCULAR ILLUSION. Paul Kirkpatrick in *American Journal of Physics*, Vol. 22, No. 7, page 493; October, 1954.

REFLECTIONS ON A GRAMOPHONE RECORD. J. B. Lott in the *Mathematical Gazette*, Vol. 47, No. 360, pages 113-118; May, 1963.

© 1989 SCIENTIFIC AMERICAN, INC

COMPUTER RECREATIONS

Catch of the day: biomorphs on Truchet tiles, served with popcorn and snails

by A. K. Dewdney

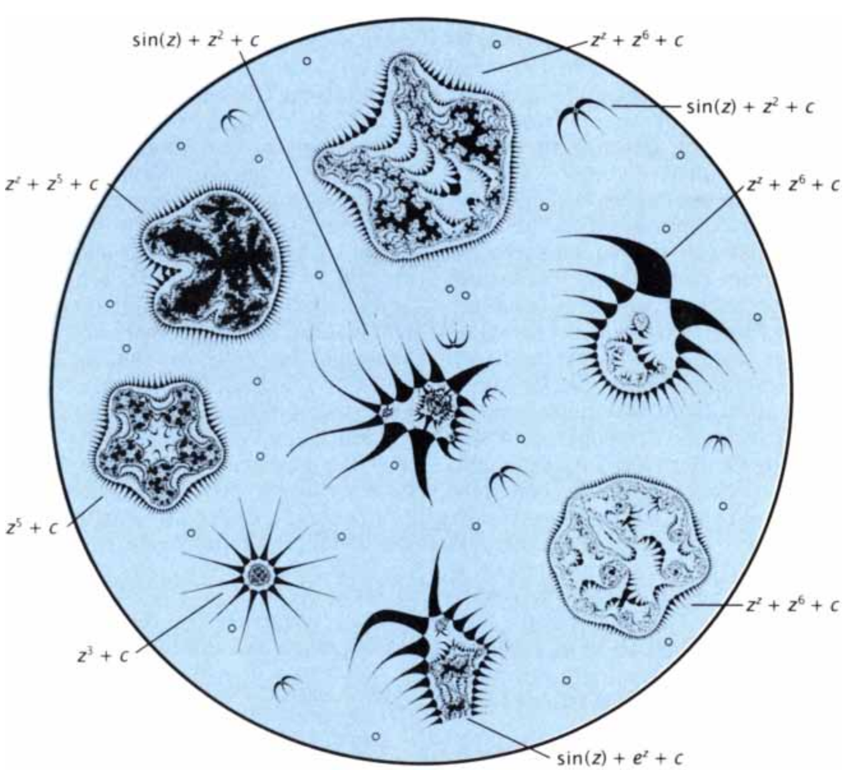
"Sometimes I consider myself a fisherman. Computer programs and ideas are my hooks, rods and reels. Computer pictures are the trophies and delicious meals."

—CLIFFORD A. PICKOVER, *Computers, Pattern, Chaos and Beauty*

On a number of occasions in recent years I have been sorely tempted to mention one or more of the strange and beautiful graphic concoctions of Clifford A. Pickover, a well-known investigator at the International Business Machines Corporation's Thomas J. Watson Research Center in Yorktown Heights, N.Y. In addition to his more serious research endeavors, he has a box full of fun programs—his "fishing tackle"—that generate chaotic but oddly aesthetic forms. I cannot resist Pickover's delectable catches any longer and have decided to describe in this column how readers can go angling for three simple but savory recreations of his: biomorphs, Truchet tilings and fractal popcorn. Readers will also get just a taste of the three-dimensional logarithmic snails captured by this soulful but sober man.

What is a biomorph? The term has appeared before in this department. In February, 1988, I described the lifelike forms created by a computer program of Richard Dawkins, a biologist at the University of Oxford. Starting with a collection of tree-shaped figures, a human operator selects one, from which the program develops another collection of figures—all varying somewhat from the one chosen. The operator selects figures in this way again and again. Eventually quite odd but distinctly organic forms may emerge.

Pickover developed his biomorphs independently and nearly simultaneously, but their method of generation and appearance are entirely different. Unlike Dawkins' biomorphs, Pickover's look distinctly microbial, which has prompted *Omni* magazine to characterize Pickover as "van Leeuwenhoek's 20th-century equivalent." (Anton van Leeuwenhoek, a 17th-century Dutch draper and lens grinder, was one of the first microscopists.)



A microscopic view of some biomorphs and their respective generating functions

Pickover's biomorphs inhabit the complex plane, site of the famed Mandelbrot set (discovered by Benoit B. Mandelbrot, a colleague of Pickover at the IBM research center). The biomorphs are produced by an abbreviated version of the process that traces the delicate fractal geometry of Julia sets, close relatives of the Mandelbrot set that were described here in November, 1987. Each biomorph is generated by multiple iterations, or repetitions, of a particular function, or sequence of mathematical operations. Each iteration takes the output of the previous operations as the input value for the next iteration.

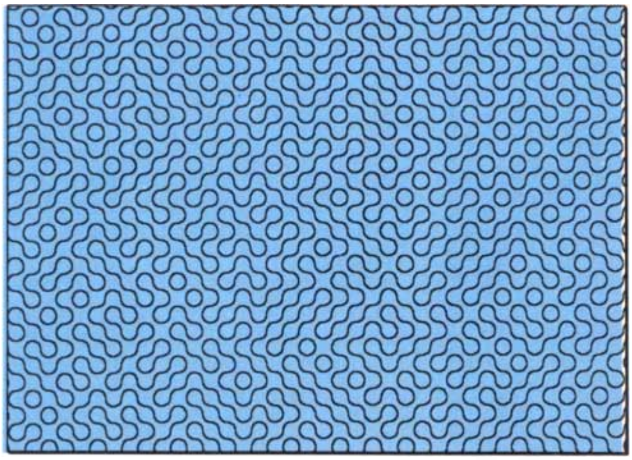
For example, the 12-spiked radiolarian biomorph in the bottom left part of the whimsical microscopic view at the left was generated by the iterated function

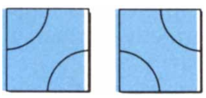
$$z_{n+1} \leftarrow z_n^3 + c.$$

An initial value for a complex-number variable (designated z_0) is raised to the third power, and a fixed complex number, c, is added to it. The same arithmetic operations are then performed on the resulting sum, z_1, to yield z_2 and so on.

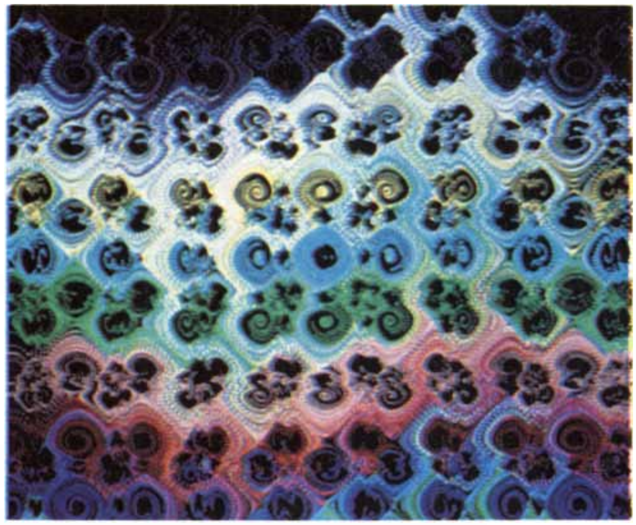
A complex number consists of two numbers of the ordinary kind; one is called the real part, and the other is called the imaginary part. The two numbers are best regarded as Cartesian coordinates on a two-dimensional plane. Traditionally a complex number is written as a sum of its two parts. For example, the complex number $3 + 5i$ has 3 as its real part and 5 as its imaginary part. (The i acts as a kind of marker to remind mathematicians

© 1989 SCIENTIFIC AMERICAN, INC





***A Truchet tile's two orientations* (left) *and a random Truchet tiling* (above)**



A cyclic system generates fractal popcorn

which part is which.) To avoid any further digression, I show in a box on page 113 how to add, multiply and cube complex numbers.

To generate a biomorph, one first needs to lay out a grid of points on a rectangle in the complex plane. The coordinates of each point constitute the real and imaginary parts of an initial value, z_0, for the iterative process outlined above. Each point is also assigned a pixel, or picture element, on the computer screen. Depending on the outcome of a simple test on the "size" of the real and imaginary parts of the final z value, the pixel is colored either black or white.

All the biomorphs pictured in the illustration on the opposite page were found in a 20-by-20 square centered on the origin of the complex plane. How many others can intrepid readers find? I shall provide the necessary viewing apparatus: a program I call BIOMORPH that follows Pickover's basic algorithm. The version I list below generates the radiolarian.

```
c ← .5 + .0i
for j ← 1 to 100
    for k ← 1 to 100
        compute z0
        z ← z0
        for n ← 1 to 10
            z ← z^3 + c
            if |real(z)| or |imag(z)|
            or |z| > 10
                then jump out of loop
        if |real(z)| or |imag(z)| < 10
          then plot (j,k) black
          else plot (j,k) white
```

The seemingly simple instruction "compute z_0" is actually not so trivial. It requires the conversion of each pair of pixel coordinates (j,k) into a complex number. This is done by dividing the length and the breadth of an area on the complex plane by the number of j values and the number of k values, respectively. The quotients then serve as the increments by which the real and imaginary parts of z_0 are systematically increased in each of the algorithm's cycles.

For example, the radiolarian is actually found in a square "window" of the complex plane delineated by the following ranges for the real and imaginary parts of z_0.

$$-1.5 < \text{real}(z_0) < 1.5$$
$$-1.5 < \text{imag}(z_0) < 1.5$$

Because j and k both run from 1 to 100, both the real and imaginary parts of z_0 are systematically increased in steps of .03. The version of BIOMORPH given above must therefore rely on statements of the type

$$\text{real}(z_0) \leftarrow -1.5 + .03j$$
$$\text{imag}(z_0) \leftarrow -1.5 + .03k$$

to compute 10,000 values of z_0, each of which is iterated and tested.

Within the innermost loop of 10 iterations, the size of z ($|z|$) as well as the size of its real and imaginary parts ($|\text{real}(z)|$ and $|\text{imag}(z)|$) is continually monitored. The size of a complex number is simply the square root of the sum of the squares of the number's real and imaginary parts, and the size of its component parts is given by their respective absolute values. (The absolute value of a number is its numerical value irrespective of sign.)

If the size of z or of its component parts ever exceeds 10, the program must promptly leave the loop—even if it is before the 10th iteration—and proceed to retest the size of z's real and imaginary parts. (When comparing the size of z with 10, it is actually simpler to compare the sum of the squares of z's real and imaginary parts with 100 than to compare the square root of the sum with 10; the result would be the same in either case.) Regardless of the number of iterations, if the size of either the real or imaginary part of the final z value is less than 10, the pixel having coordinates (j,k) is colored black. Otherwise it is colored white.

In most personal computers the pixels whose coordinates range between 1 and 100 will lie in one corner of the screen. To center the image, the initial and final values of j and k may have to be altered. For example, j and k may have to run from 50 to 150 instead of from 1 to 100.

All other details about BIOMORPH will have to be worked out by the adventuresome. I shall only remind readers that z, z_0 and c are all complex numbers and must be added and multiplied as such. For that same reason every assignment statement in the algorithm involving z, z_0 and c represents two assignment statements in ordinary computer languages—one for the numbers' real parts and one for their imaginary parts.

How are the other creatures made visible through Pickover's imaginary microscope produced? Actually, BIOMORPH can produce them as well. One only has to substitute a different iterative function at the heart of the algo-

© 1989 SCIENTIFIC AMERICAN, INC

A logarithmic function yields a snail shell

rithm. In other words, one merely replaces $z^3 + c$ in the above program by another function.

Those who write the BIOMORPH program may well want to fish for other creatures that inhabit the central 20-by-20 area of the complex plane. The simplest search would entail scanning the entire area; one can then zoom in to inspect the graphic details of whatever creatures the preliminary search turns up. Any small value of c will normally serve as bait.

Some alternative biomorph-generating functions make use of a strange power of z, z^z; others rely on trigonometric operations. Pickover has kindly agreed to send those who ask a short primer on the mysteries of complex-number trigonometry and self-powered complex numbers. Armed with that knowledge, one can write versions of BIOMORPH that yield the protozoan morphologies shown.

Biomorphs began for Pickover as a bug in a program intended to probe the fractal properties of various formulas. He accidentally used an OR instead of an AND in the conditional test for the size of z's real and imaginary parts. That unintentional change colors a great many more pixels black than would otherwise be the case. The cilia that project from the biomorphs consist solidly of such pixels.

Although they were serendipitously created, the biomorphs seem to have taken on a life of their own. As Pickover puts it: "In some sense the mathematical creatures exist. These objects inhabit the complex plane, though they resemble microscopic organisms that we could easily imagine flourishing in a drop of pond water." In what sea might we fish for more advanced forms of life? Pickover also finds it meaningful that the complexity of both natural and artificial organisms results from the repeated application of simple dynamic rules.

Pickover's current work at the research center revolves around new ways of rendering complex data meaningful at a glance. For example, ordinarily it is quite difficult to detect regularity or randomness in a page full of 0's and 1's, which is a fundamental way to depict bits of data. Yet Pickover can make the data's degree of randomness clear by means of a simple graphic device based on Truchet tiles, which get their name from the 18th-century French monk and polymath P. Sébastien Truchet. Pickover has modified Truchet's original tiles to consist of two quarter-circles in a square, each centered at opposite corners and intersecting two of the square's sides at their midpoints. The resulting tile [*see illustration at left on preceding page*] exhibits only two distinct orientations.

If one paves a flat area with Pickover's Truchet tiles, strange sinuous curves appear. There are no loose ends, except at the edges of the area. Each curve is guaranteed to be continuous because the midpoints of the sides of adjacent squares (which is where the quarter-circles begin and end) abut.

Any array of 0's and 1's can be converted into a Truchet tiling merely by arranging the individual tiles in the same order as the binary digits, orienting a tile one way wherever there is a 0 and the other way wherever there is a 1. If the data bits are random, Pickover maintains, one discerns no particular pattern in the tiling: the curves loop and meander in a perfectly confusing (albeit fascinating) way.

Introduce a slight element of regularity in the data, however, and definite patterns of curves are perceived. For example, if one generates an array of bits in each row of which a 0 is slightly more likely to be followed by another 0 than a 1 and a 1 is slightly more likely to be followed by another 1 than a 0, the curves in the corresponding tiling would show a distinct diagonal trend. Readers may enjoy discovering for themselves why this is so.

Actually, I prefer the tiling patterns created by random bits, such as the one shown in the illustration. They suggest a host of recreational pastimes. Are they mazes, for instance? Try to find a channel that leads from the top of the tiling all the way to the bottom or from one side to the other. In the course of tracing channels, one often comes across "islands"—curves that rejoin themselves. Any curve that does not strike the edge of the square must of course come back to itself sooner or later. Some of these curves are small circles; some are dumbbell-shaped, and some are more complicated still. Readers are encouraged to classify the other types of closed curves. What are their average numbers per tile?

Writing a program that turns bits into Truchet tilings is easy. Assuming that one has an array of binary digits to begin with, one merely sets up a double loop that scans each element of the array and draws the corresponding Truchet tile at the appropriate place on the computer screen. This much explanation will have to suffice so that I can turn to another Pickover graphic recreation: fractal popcorn.

Pickover's popcorn stands out, curiously tactile, in the illustration at the right on the preceding page. It is ac-

© 1989 SCIENTIFIC AMERICAN, INC

tually a tracery generated by the solution of a pair of iterated differential equations in discrete form. The x coordinate of a solution point is computed by subtracting the value of a function of y from the x coordinate of the previously calculated solution. Likewise the y coordinate of a solution point is computed by subtracting the value of a function of x from the y coordinate of the previously calculated solution. Such a set of equations is called a cyclic system. Cyclic systems based on trigonometric functions seem to be a particularly rich source of spectacular graphics.

The crossed-variable iteration process in the program POPCORN is described by the pair of equations:

$$x_{n+1} = x_n - h\sin(y_n + \tan(3y_n))$$
$$y_{n+1} = y_n - h\sin(x_n + \tan(3x_n))$$

A small constant, h, multiplies the trigonometric functions so that a new value of x or y is never far from the previous one, regardless of how long the iteration continues. Pickover gives h a value of .05 in his program.

As the following algorithmic outline indicates, POPCORN iterates the formulas 50 times for each of 2,500 initial-pair values (x_0, y_0), each time plotting a pixel on the computer screen.

```
for j ← 1 to 50
   for k ← 1 to 50
      x0 ← -6 + .24j
      y0 ← -6 + .24k
      x ← x0
      y ← y0
        for n ← 1 to 50
           xx ← x - hsin(y + tan(3y))
           yy ← y - hsin(x + tan(3x))
           x ← xx
           y ← yy
           compute jp and kp
           plot (jp,kp)
```

(The variables xx and yy are temporary variables that serve to keep track of the current values of x and y within an iteration cycle.)

The initial values for both coordinates are chosen from points within a square area of 12-by-12 units centered on the origin of the x-y plane. In other words, initial values of x are between −6 and +6, as are the initial values of y. Like BIOMORPH, POPCORN relies on a nested loop structure to compute a set of initial values. In POPCORN, however, both loops run from 1 to 50.

Finally, the points specified by the 50 (x,y) pairs generated for each initial pair (x_0, y_0) must be plotted as pixels on the screen. That is done by a simple formula that scales and shifts x and y to jp and kp.

$$jp \leftarrow 4.166x + 25$$
$$kp \leftarrow 4.166y + 25$$

Pickover's fractal popcorn therefore consists entirely of pixels whose coordinates are given by the iterations. Readers are free to zoom in on individual pieces of popcorn to discover quite beautiful and completely unexpected detail. They can also "dye" the popcorn, as Pickover does, by plotting the pixels in a different color each time the inner loop begins anew.

Pickover's graphics have caught the public eye on more than one occasion. His work has been discussed not only in various magazine articles but also in television shows; it has even been shown at exhibitions in Switzerland, Japan and at the Computer Museum in Boston.

I shall end this column by drawing the reader's attention to one of Pickover's latest creations: a beautiful snail shell, complete with stunningly realistic coloring and highlights. The image, shown in the illustration on the opposite page, is constructed from a sequence of spheres whose centers track a gentle, logarithmic spiral. The highlights are generated by a ray-tracing technique that mimics various kinds of light reflections.

To obtain the guidelines on complex arithmetic or to receive a special bibliography of Pickover's papers, readers may write to Pickover at the IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y. 10598.

The subject of this department in March was vandalware: malicious programs that replicate themselves and do damage to benign software. The plague of such computer "viruses" and "worms" continues. Rumors of a new virus that attacks spread-sheet programs have been circulating for more than a year, but now they have been confirmed. Virus expert Harold J. Highland, who edits the journal *Computers and Security,* calls the new infection a macrovirus, because it consists of special commands called macros.

Reports of the macrovirus are difficult to track down, apparently because it has struck corporations that are reluctant to release details to the public. One version of the virus appears to be spreading on Lotus 1-2-3 software. Every time a user runs an infected program, the virus calls up a work sheet from memory at random, alters the numerical content of a single cell by a small amount and puts the altered work sheet back in memory. The damage the virus can do is potentially great, since small changes in data can go unnoticed for months, significantly affecting the results of the spreadsheet calculations.

In spite of such evidence, Tom Pittman of Spreckels, Calif., thinks the March column was an overreaction on my part. According to Pittman, one must be careful to distinguish between useful and destructive viruses. The difference may be semantic. There are certainly useful programs that lead a viral or even wormlike existence. They have jobs that range from file compression to input scanning, but they are not normally called viruses or worms.

Edwin B. Heinlein of Mill Valley, Calif., would press for the establishment of a Center for Computer Disease Control modeled after the Centers for (Human) Disease Control in Atlanta. Such a center would be part of a broader effort for studying and categorizing viruses and for drawing up new laws pertaining to the reporting of incidents and the punishment of offenders. The research undertaken in such an effort might also yield new filters and virus-detection schemes.

Two complex numbers are added by summing their respective real and imaginary parts separately. In other words, the sum of the complex numbers represented by $a + bi$ and $c + di$ is $(a + c) + (b + d)i$.

The product of the two numbers is slightly more complicated. It is $(ac - bd) + (ad + bc)i$. That formula can be applied in order to derive a formula for the square of $a + bi$, which is $(a^2 - b^2) + 2abi$. The square's real part is $(a^2 - b^2)$ and its imaginary part is $2ab$.

One further application of the multiplication formula yields another formula, one for the third power of $a + bi$. The real part of $(a + bi)^3$ is $a(a^2 - 3b^2)$, and the imaginary part is $b(3a^2 - b^2)$.

Complex-number arithmetic

FURTHER READING

BIOMORPHS: COMPUTER DISPLAYS OF BIOLOGICAL FORMS GENERATED FROM MATHEMATICAL FEEDBACK LOOPS. Clifford A. Pickover in *Computer Graphics Forum,* Vol. 5, pages 313–316; 1986.

MATHEMATICS AND BEAUTY: TIME-DISCRETE PHASE PLANES ASSOCIATED WITH THE CYCLIC SYSTEM. Clifford A. Pickover in *Computers and Graphics,* Vol. 11, No. 2, pages 217–226; 1987.

© 1989 SCIENTIFIC AMERICAN, INC

BOOKS

Cats and people, algorithms and Crays, spicules and wheels, goonch and neem tree

by Philip Morrison

The Domestic Cat: The Biology of its Behavior, edited by Dennis C. Turner and Patrick Bateson. Cambridge University Press ($59.50; paperbound, $22.95).

Cats have dwelt amongst us for some 4,000 or 5,000 cat generations—enough to select out a new subspecies even though there was not always a systematic control over breeding. The cat fancy has maintained that control tightly for a few hundred years at most. Here in the U.S. there are about 50 million cats in and about our homes, barns and fields; the cat is fast becoming the favorite pet of a busy urban population in most countries of the industrialized world.

Cat behavior is the topic of plenty of many widely read books, most of them anecdotal and personal, their insights more artful than scientific. This volume sets out the status of our knowledge—and our ignorance—as documented for many cats in many places, the behavioral research of a score of experts from Switzerland to New Zealand. They are professional students of pet, farm and feral cats, bringing to bear all the techniques and expertise of contemporary ethology, from cat-by-cat and moment-by-moment recording to radio location. The chapters have been worked into an overall account, from the individual development of kittens and their mothers to cat societies large and small (with their classes of membership) and on to the hunting behavior of a deft predatory carnivore. There is some untangling of the well-woven skeins that have come to link cats and people: the present of a rubber ball, the generous cat gift of a half-dead mouse or the friendly rubbing of cat cheek on human nose. Those events and more are shown in attractive and authentic line drawings by Melissa Bateson.

The dog has shared the fireside with us since cave days. The circumstances of the imperfect domestication of the cat remain a little speculative. It was much more recent, maybe later than 3000 B.C., with the tabby-coated wild North African subspecies of *Felis silvestris libyca* as the chief ancestor of the present somewhat hybridized domestic cats. That first subspecies is clearly more docile than its intractable northern European counterpart. An Egyptian tomb of 1900 B.C. holds the bones of 17 cats, along with little pots for offered milk. By 1600 B.C. cats were fully domesticated in Egypt, often painted in family and hunting contexts. They were adored pets and cult objects for five centuries before Herodotus wrote of the cat goddess and her priesthood and of the great temple that was visited in spring by 700,000 festive and bawdy pilgrims. Cats came still later to Rome and then to India and the Far East.

Both in Christian Europe and in medieval Japan, cats have encountered times of popular love and of fierce hate. Hatred of the cat is associated with witchcraft and demonic possession, an imputation whose roots certainly include a negative perception of active female sexuality and the "sexual insecurities of . . . males." Popular as cats are today, there is evident public ambivalence (on the hard evidence of polls) about the cat who walks by herself in the wet, wild woods. Dogs, so loyal and so deferential, do not share such antagonism. All the same, a 10th of the cat and dog populations alike are put to death every year in the U.S. as unwanted and expendable.

All cats hunt, hungry or not, when they can. Mothers bring dead prey to their kittens very early, and then during weaning they release live prey near the kittens, letting them learn how to capture and kill. Kittens follow the mother's choice, preying lifelong mostly on the same strains of mouse or rat. It is plain that there are particular developmental stages during which the kitten is open to a variety of lessons through trial and play; once each stage is passed, the kitten is quite resistant to change in the learned behavior. Preference for social companions, in particular human beings, is largely fixed in the time between two and seven weeks of age.

The diet of cats, as measured both by stomach contents and by animals brought home, shows that among the small mammals cats prefer voles, then young rabbits and hares, then mice and rats, with insectivores last. Birds are less important items of diet on all continents; insects and reptiles are eaten as well. Cats given milk as a supplement can keep buildings free of Norway rats within a distance of about 50 meters from the place where the cats dwell, "once existing infestations [are] eliminated by other means." Because cats prey mainly on young rats of under 100 grams, a full rat population will survive the cat onslaught a long time. This is the supposed basis for early domestication: it is argued that the cats "domesticated themselves" by hunting in the pest-ridden Egyptian granaries, to be lured and rewarded by observant and grateful farmers. Yet the scenario ignores the Egyptians' fondness for taming animals, demonstrated with monkeys, hyenas, mongooses, crocodiles, lions and many horned beasts. They would never have overlooked cats, even if cats could not pay their way by catching mice.

Cat society is wonderfully reported here, from solitary cats (who yet from time to time hear and scent others) to bands of feral cats on uninhabited islands, where they live by predation on seabird nesting colonies. Here are the scandalous affairs of the cats of the Roman ruins (observable only by day), with scores of cats to a band, fed daily by cat lovers, and tales of even larger groups in a Japanese fishing village living on the rich dumps of fishing waste. There are careful studies of numbers and kinds of interaction, pair by pair in colonies of various sizes.

The matter is far from simple, but it is hard to avoid the conclusion that all cat society is based on the "social matriline," the group consisting of one female and the generations of her offspring. Stable cat society is a set of female kin groups, around which loosely circle some adult males, young and old, who come and go. Food supply fixes density, whether it is the regular feeding by Swedish nonfarm rural households or the abundant young rabbits taken as prey out in the wild grasslands of Australia. The best "reproductive strategy" for a species given to multiple mating is not easy to

© 1989 SCIENTIFIC AMERICAN, INC

fix, despite considerable effort. There are no such studies yet of the wild ancestral subspecies, but here there is an echo of the pride of lions, if not the dark lair of the tiger.

The genetics of behavior is surprisingly absent from this account. There is mention of the deafness of blue-eyed white cats, whose females are often unusually timid. Siamese cats have a serious wiring error in the circuitry between the retina and a junction box and processor along the optic nerve. The mix-up is believed to be the consequence of a single enzyme deficiency. Somehow the Siamese can compensate well enough to see almost normally. The special behavior patterns of selected breeds of cat are mentioned here only in passing.

The volume is a fine example of behavior study; it is a proper study for humans, for surely we share the cat's mammalian heritage. The plasticity of cat behavior, built of a modest number of modules that can be arranged to suit the widely varied environments we have taken them into, is worth our keen attention. Our primate-derived and culture-overlain nature is certainly no less adaptable; we shape our tame and wild selves more freely even than we have formed the cats, if within still dimly seen constraints of our own evolutionary legacy. One gathers that not all the sociobiologist authors of this lively and readable book would agree.

IDEAS AND INFORMATION: MANAGING IN A HIGH-TECH WORLD, by Arno Penzias. W. W. Norton & Company ($17.95). **SUPERCOMPUTERS AND THEIR USE,** by Christopher Lazou. Revised edition. Oxford University Press ($55; paperbound, $27.50).

How would you find all the anagrams in a piece of text—say, in "Can cane sugar lead to a good deal of acne?" Matching the letters of the first word against those in each of the following words is a brute-force method, fine for a line or even a page, hardly for a bookful. Better to spend your comparison effort only with words of matching length, still better to sort the words by length before comparison, better yet to alphabetize the words first in order to avoid repetitive comparisons. A winning scheme is rather to alphabetize the letters within each word and then to arrange those respellings in alphabetical order: "a, acen, acen, acn, adel, adel. . . ." The two anagram pairs now match easily. That is an instance of a sensible algorithm, or sequence of elementary steps required to gain an end: the intellectual module of computer programming. "In my view, the ability to find such creative solutions comes closer" to computer literacy than memorizing keyboard commands, Penzias concludes.

This demystifying, modestly predictive brief book in a personal and conversational tone, innocent of jargon, equations and even diagrams (one simple logical circuit does appear), is the general reader's highroad into the modern world of computers and their consequences. The author is a famed radio astronomer now become the reflective research head of the AT&T Bell Laboratories, one of the world's oldest and most successful industrial research establishments.

Sound-recording companies do not keep analog libraries of sweet sounds any more. Instead they sample the amplified signal of any performance every few microseconds and measure its amplitude, recording all "the 'wiggles' in numerical form." A cookie cutter is a homely example of analog information; an old cutter will form cookies in its own marred image to share its dents and nicks. But the digital text of an old cookbook recipe, like a digital compact disk, will continue to prescribe cookies just as good as they ever were until the creases and stains lie so heavily on the page that whole words become uncertain. Digital information is resistant to noise, although not totally immune.

So runs Penzias's account, with essays on numbers, rules, technology, ideas, the implications of computing; on it goes to the human element of the new technology, the hierarchies of good and bad organizations, and finally a small biography of Theodore N. Vail, the executive who created the modern Bell system between the times of Teddy Roosevelt and Warren G. Harding. Along the way one hears of the formidable chess-playing computer Belle, the pseudopsychiatrist ELIZA and the rise of optical-fiber communications. The virtue of rough estimates, the inordinate increase of permutation possibilities with the number of elements, and the productive integration of diverse individual expertise are strong motifs here, with a neat summary of the rise of Apple from surging innovator to a "big, boring" and brilliantly successful corporation. This is not surprising from the viewpoint of a man at home within AT&T, but the argument is in place for dissenting readers to reflect on.

One cautionary anecdote of many must suffice. At Holmdel, N.J., the big radio horn—the one with which the author first detected the afterglow of the big bang—was never used for transmission. Yet even during idle seasons a neighbor complained that its presence ruined his TV reception. In vain the Bell Labs people explained how that could not be, and finally the engineers accepted his despairing challenge to come see for themselves. It was all true; the innocent viewer, not the experts, was right. The big immobile structure reflected one particular TV signal only at his receiving location, to cancel and break up the good picture once enjoyed! Offered as an example of the open-minded tests required by the method of science, the tale echoes with wider implications—natural enough given the author's own openness and good humor.

The second of these books is by no means meant for the computer-innocent. It is a tour de force of savvy epitomizing and appraisal. The volume reminds a reader very much of the genre of books that picture and compare the warships or the aircraft of the world. Neither the technically uninformed nor the directly engaged specialist can use such books very well; rather they fascinate and support a culture of the curious, the critics and those who want to come to know not so much the trees as the forest. To bring out the aptness of this comparison, it is enough to look at one page of this lively book. There side by side are scaled line drawings, one from the top and one from the side, of Cray-2 and Cray-3. Cray-2, a sculptural beauty with a C-shaped plan five feet across, first was flown at NASA Ames in late 1985. Its faster, more powerful brother, a yard-high octagon, was in prototype last year at Chippewa Falls, Wis., where Lazou admired its robot assemblers working in their crowded clean room, its new and faster gallium arsenide chips and its intense miniaturization. Cray-3 should reach its first users within a year, giving them five or 10 times the speed and memory of Cray-2.

Supercomputers are as old as digital computing itself: the name is simply given to the fastest and most powerful general-purpose computers extant at a given time. They are primarily "number crunchers," 100 times faster than commonplace big mainframes. They are thought of as scientific computers even though they are put into play far from the research lab or design department—for decryption, communications traffic analysis and the moviemaker's digital special effects. Perhaps 70 supercomputers, half of the population, were government-operat-

© 1989 SCIENTIFIC AMERICAN, INC

ed in 1985. In 1990 there will be many more in industrial and university use: predicting fluid flow for aircraft, automobiles, reactors and plasmas; solving many-body problems; doing 3-D seismic mapping; and analyzing complex structures, ocean currents and real-time weather. The world's supercomputer census should rise by the early 1990's to 450 machines. Six corporations in three continents are now in the market; they expect as much as $20 million for each machine.

The main text here is a few years old; its account of the history, the chief players and the main technologies, issues and applications is still fully useful. The final fourth of the revised version is a very knowing update to the spring of 1988. There is a brief appendix on the features of modern FORTRAN for array processing.

The ticks of the supercomputer's clock are fast: it beats every couple of nanoseconds. Operation is concurrent as well as sequential; parallel flows of data and operations stream down scores of channels all at once to multiple central-processing units, 16 of them in the new Cray. The fast machines are physically small, because subunits must communicate and long paths carry two-way signals slowly. These dense, active machines necessarily run hot: 150 kilowatts in well under a cubic meter of intricate electronics is comparable to the power of an automobile engine, one made not of aluminum and steel but of the most delicate materials. Cooling is insistent and even determining; a fluorocarbon coolant pours through everywhere. Memory is vast, a couple of gigawords of 64 bits each.

What comes clearest is that supercomputing success is a matter of many-sided balance. Multiple parallel processing is needed (it has been used for a generation at a low level of multiplicity), but so too is swift one-by-one operation, capable of executing a dozen floating-point logical steps between clock ticks. Interlinking many parallel streams takes time, exactly as waiting for step after step takes time. A chain of events is no faster than its slowest links. Ingenious new "massively parallel" architectures, with many small, slow central-processing units of modest power interconnected (as if along the many edges of a multidimensional hypercube, for example), promise good cheap specialized machines—minisupercomputers, now in the making by a hundred smaller firms. They are not yet the top of the line; someday, perhaps. Rumor has one famous designer now quietly at work on a million-processor machine. Fiber-optic speed beckons: imagine a cluster of 16 Cray-3's serving a superfast shared central memory, all optically linked.

Somewhere ahead convergence lies. More multiple but still fast-working conventional architecture will encounter speeded-up massively parallel arrangements somewhere on this side of infinite speed and of the new millennium. Software, now based on 30-year-old FORTRAN (to be sure in its new standard 8x avatar), is growing in richness. Out there on the expanding horizon the giant shadow of IBM can be made out clearly: Cray-4 will one day be matched against a new IBM counterpart and possibly some multimegawatt monster of multiple processing as well, called IBM TF-1 for its hoped-for teraflops. The subculture is in luxuriant blossom.

LIFE'S DEVICES: THE PHYSICAL WORLD OF ANIMALS AND PLANTS, by Steven Vogel. Illustrated by Rosemary Anne Calvert. Princeton University Press ($49.50; paperbound, $17.95).

Security is not always to the strong. The bull kelp in its cold, wave-tossed forest is held to the bottom on a long, tensed stem, but the strength of that stem is "orders of magnitude below" tendons or cuticle. Only its tensile-energy storage is at a respectable magnitude; the rubbery stuff could recover from extensions of almost 50 percent. It endures by giving way for a while, since it is a rare single wave that lasts long enough to reach the breaking strain.

Steven Vogel is a Duke zoologist whose special study is biomechanics, the engineering of organisms mostly at a macroscopic scale. His own "infantile prejudice" in favor of fluids, relict of the universal nursery-school delight in pouring water, led him to an earlier volume that was narrower and deeper than this, if still a wonderfully broad introduction to the physics of life in moving fluids. Here he offers readers a more basic and wider look at life for the sheer delight of winkling out from its diversity the "constraints and opportunities" directing that patient engineer, evolution. He writes with unusual recognition of the needs of the inexpert reader, in a personal and lighthearted style, seasoned perhaps a little too strongly with punning and word play.

It would not be easy to cite a better brief introduction to literacy in physics than the first four chapters of this volume, along with the notes on numbers at the end of the book. Graphs and tables, with very few equations, lead a novice through the needed geometry and mechanical understanding, mainly the big ideas of conservation, variables and ratios, and dimensions. They are all given concreteness by illustrative tables and nifty examples. It is not all pat, either; there are plenty of surprises. "Just as you can guess the size of the animal from a picture of its skeleton," explicit with bony drawings of cat and elephant at equal size, "you can guess the size of an animal from a photomicrograph of one cell of its liver." Animal cells are pretty much all of a size, but their internal biochemical plant adapts to overall metabolic needs; Vogel does not mention mitochondria, although later he does enter the microworld for muscle fibers and microtubules.

The remainder of the book builds chapter by chapter on that strong base of physics. Comparative biomechanics is pursued through fluids, materials, structures and support on to mobility and a wise little final essay on energy. It is not that there are many direct lessons for human engineers in the works of the natural one, but where else is there any technology to compare with what we have? "Ships don't wiggle their sterns, cars aren't... equine, and airplanes don't flap," all for very practical reasons. Even if sponge-spicule networks superficially evoke the tensegrity nets of Kenneth Snelson and R. Buckminster Fuller, sponges are not at all the "Fuller-on-earth" that Vogel once expected. Sponge networks are highly deformable, almost the opposite of Fuller's cleverly won and wiry stiffness. (Vogel reminds us that the bath sponge is an aberrant form, almost without spicules—all soft stuff.) Some of the grand tents of the designer Frei Otto lie closer to many pressurized structures of life.

Wheels are a second example. Life at macroscale does not roll, whereas so many of our machines do, inside and out. Some evolutionary authors have dilated on how hard it is to evolve organic free-spinning wheels, but Vogel cites his "fractious friend" Michael LaBarbera on how much the utility of wheels depends on roads: "The general superiority of wheels as a mode of transport is false." Consider the narrow, winding streets of the Casbah, welcoming camels, donkeys and porters but not carts or cars.

A reader from New Mexico might be forgiven some skepticism toward both rationalizations. See the vision of some large prairie plant blowing across the wide, undulating land in a

© 1989 SCIENTIFIC AMERICAN, INC

steady wind, rolling swiftly on a few wheels made of its dried and much-branched outer twigs, now loosely held by the sturdy dried stem that once fed them. The dried tumbleweed is visibly all wheel; evolution is surely capable of such a design, given need. When we see life on other planets, look out for wheels. Meanwhile, this delightful volume leads to the only comparable technology around, almost everywhere instructive, though hardly the same.

ENCYCLOPEDIA OF INDIAN NATURAL HISTORY, R. E. Hawkins, General Editor, Doris Norden and Bittu Sahgal, Illustrations Editors. Published on behalf of the Bombay Natural History Society by Oxford University Press ($49.95).

The grand subcontinent that contains India and half a dozen neighboring nations is home to wildlife whose scope and variety would dazzle us, but for the fact that the glow of the diverse human cultures is so bright. For the people of the region, though, the natural legacy is pervasive: the animals and plants of snowy mountains and hot desert, of deep forest and wide shores, support their lives and inform all their expression. What would be left of our image of that world without tiger and peafowl, falcon and cobra, banyan and lotus?

A century back "eight gentlemen of Bombay" formed the Natural History Society there; it has grown to national stature, a leader not only in field research and collection but in the urgent movement for nature conservation. Convinced that broad public understanding is essential to the preservation of wildness, the society decided that its centenary celebration should take the form of this book, the society's most ambitious publication: a reference work in language open to general reader and young student and a reliable first resort for the expert. This thick volume presents 550 articles, from Accentors to Zoological Survey of India, with plenty of visual matter, including a very well made endpaper map, 40 full-page plates and an ingenious set of decorative initials, each one bearing some half-dozen sketches of animals and plants with names that fit the letter in English, Hindi or taxonomical Latin.

The articles, some concise and some extended, are by many authors. They center on regional natural history. But what would any curious reader think of an encyclopedia so provincial that it did not discuss chlorophyll or the giant panda? Both are here, with many more in the same spirit. A review does better to extract the particular. The work is clear and usually satisfying at an introductory level. The language is quite free of confusing usages, and the content is free of those folkloric whoppers that beset even careful writers in this vein.

Game fish are taken widely in India; the largest is the goonch, a freshwater catfish that may run to more than 500 honest pounds in weight. Tornadoes are rare; one in Delhi in 1978 was viewed as exceptional. But the tracks of hurricanes (colorlessly called severe cyclones) enter the peninsula from one coast or the other a few times each year, mostly before and after the monsoon period. Severe cyclones in the Bay of Bengal are twice as frequent as those on the Arabian Sea and are disastrous to the Bengal lowlands. The lofty poonspar tree is a glory of the wet evergreen forests of the Western Ghats. For hundreds of years it was felled to provide masts for seafaring dhows; "the price paid was the number of rupees laid edge to edge along the length," which was often 45 meters.

In the Deccan lava flows, a million cubic kilometers of basalt, there is a famous pile of hexagonal columns no less remarkable than those of the Hebrides or of Wyoming. Elsewhere in those wide, steplike formations are the "railway diamonds," geodes filled with gemmy quartz found abundantly during the boring of the railway tunnels near Pune. The commonest garden spider in India weaves an orb web, raises its center into a dome, builds an open-meshed deck below the top of the dome and waits there upside down for prey; often one finds several such spiders side by side in a thorny bush. Two hundred thousand Olive Ridley sea turtles nest each year on one beach on the Bay of Bengal; in good years many thousands of pink flamingos nest on the salt flats of the Great Rann of Kutch; every day in the parks of New Delhi, and indeed all over the north, the thrushlike birds they call in Hindi the "seven sisters" quietly hop along the leaf-strewn ground, always in amiable little parties.

For such other renowned Indian topics as falconry, elephants, tigers, the greater rhino, the king cobra ("most intelligent of all snakes" and the only snake that nests) or the wonderful neem tree, whose beautiful deep-green antimalarial leaves are in universal daily use as good for the gums—for all these you will have to explore this unusually attractive book on your own.

SCIENTIFIC AMERICAN

CORRESPONDENCE

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

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.50 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/89. $1.50 + 0.00.

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

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

Address subscription correspondence to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 3187, Harlan, IA. 51593. Telephone inquiries: 1-800-333-1199, U.S. only; other 515-247-7631/32. The date of the last issue on subscriptions appears on each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, a mailing label of a recent issue) as well as the new one.

Name ______________________

New Address ______________________

Street ______________________

City ______________________

State and ZIP ______________________

Old Address ______________________

Street ______________________

City ______________________

State and ZIP ______________________

© 1989 SCIENTIFIC AMERICAN, INC

ESSAY

From Descartes to neural networks

by Patricia Smith Churchland

Reflex behavior, Descartes reasoned, could be explained in terms of the physical connections of nerves and muscles. But for nonreflexive behavior, behavior produced by intelligence and planning, he could not imagine a physical explanation. In his famous summation "I think, therefore I am," the "I" was considered a nonphysical thing, a soul existing independently of the physical brain but in some kind of causal interaction with it. The Cartesian idea of two very different substances—the mental and the physical—has enduring appeal. The subjective phenomena of which we are introspectively aware when we see or think seem to be entirely different from the physical activity of brain cells.

Yet how things seem is often not how science discovers them to be. The earth does in fact move even though it seems not to; "solid" matter is in fact mostly empty space; living things were not created but evolved from simpler structures. Descartes and introspection notwithstanding, those of us who are physicalists expect that by discovering the nature of the brain and the principles governing how it works we can understand perception, learning and other "mental" functions in neurobiological terms.

How fares the physicalist program? Within this decade progress in neuroscience has been spectacular, and a great deal is now known about the properties of neurons and about the complex molecules that affect the responses of neurons. At the same time experimental psychology and clinical neurology have yielded behavioral data that reveal much about the scope and character of such psychological capacities as visual perception and memory. These data are essential if neuroscientists are to know exactly what the functions are for which they are seeking mechanisms.

If we already understand much about the nature of cells and about the general character of psychological capacities, is that not sufficient to explain how we learn, see or talk? No: necessary these data surely are, but sufficient, alas, they cannot be. The reason is that the brain is a kind of computer, and to understand how the brain works we need also to understand the computational principles that nervous systems employ. To be sure, neurons are the basic units, but they interconnect to form networks and systems. If we are to understand how the brain enables us to see and learn, we must understand how networks of cells interact to represent, transform and store information.

The hard part is to figure out what kind of computer the brain is. The problem would be easier if the brain resembled the familiar von Neumann computers, with their serial, digital architecture, unmodifiable connections and memory banks. The resemblance is very feeble, however; the brain seems to be a computer with a radically different style. For example, the brain changes as it learns, it appears to store and process information in the same places, its elements are analog rather than digital and it is comparatively fault-tolerant. Most obviously, the brain is a parallel machine, in which many interactions occur at the same time in many different channels. Moreover, natural selection being what it is, there is a premium on solutions that are fast and approximate rather than slow but exact.

The dramatic conceptual breakthrough has been the invention of computer models that to a first approximation are brainlike. Neural-network models (also called connectionist models or parallel distributed-processing models) attempt to capture, at some appropriate level of abstraction, the computational principles governing networks of neurons in nervous systems. Typically the models have neuronlike units, axonlike lines connecting the units, and modifiable synapselike weights on the connection lines.

A major discovery has been that these model networks can learn. Rather like organisms with a nervous system, they can extract commonalities from examples and generalize to new cases. The key to their learning is that their synapselike weights can be modified incrementally so that the answer the network gives to a question gets closer and closer to the correct answer. How do the weights know in what direction and by how much to change? Various algorithms have been devised to adjust the weights of synapses—as a function of error on the previous input-output trial, for example.

By means of such algorithms a network can be trained—as opposed to simply being programmed—to perform surprisingly complex tasks. For example, it can be trained to distinguish the sonar echoes of rocks from those of metal objects. It is technologically important that simple network systems can thus learn to solve problems of such complexity; the finding is also illuminating theoretically, because real nervous systems too must learn by using various strategies for the modification of synapses. Just what algorithms the brain actually uses for synapse modification is not known, but the hope is that convergent research from neuroscience and from network modeling will be able to discover them.

Other seminal ideas emerging from neural-network modeling have provided insight into what the brain could be doing. The general problem of the nature of computations and representations in nervous systems is now more approachable because cognitive representations in model networks simply *are* patterns of activity across a large population of units; computations are synapse-mediated transformations from one pattern to another. All of this means that representations must typically be distributed across large neuronal populations rather than being assigned to individual "grandmother" cells. Motor control is likewise distributed rather than emanating from "command" neurons.

Although the modeling of nervous systems is still in its infancy and we do not yet have any model that exactly explains how we see or learn, there is a gathering conviction that current lines of research, like the network models themselves, are converging on answers. Progress so far does provoke educated speculation about the neurobiological basis of our mental lives. Descartes' was a 17th-century vision: pre-Darwin, pre-Turing and pre-neuroscience. Informed by neuroscience and computer science, we can modernize his vision and begin to discern the shape of a new theory about the nature of the mind—of what it is for the physical brain to see, learn, and understand itself; of what it is to be a human being.

PATRICIA SMITH CHURCHLAND, professor of philosophy at the University of California, San Diego, is the author of *Neurophilosophy: Toward a Unified Science of the Mind-Brain.*

© 1989 SCIENTIFIC AMERICAN, INC

ELEK-TEK® 10th year 1979 1989

Great Products at Great Prices from a Great Company

ELEK-TEK®

HUGE DISCOUNTS ON TOSHIBA EPSON OKIDATA HEWLETT PACKARD XEROX / Diablo

PRINTERS & ACCESSORIES

LX-800 $179

OTHER EPSON PRINTERS

LX 800 180 CPS **$179**
FX 850 264 CPS .... **CALL**
FX 1050 264 CPS .... **CALL**
DFX 5000 533 CPS .... **CALL**
LQ 500 180 CPS **275**
LQ 850 264 CPS .... **CALL**
LQ 950 264 CPS .... **CALL**
LQ 1050 264 CPS .... **CALL**
LQ 2550 NLQ 324 CPS .... **CALL**

TOSHIBA PRINTERS

P321 SL .... **CALL**
P321 SLC (color) .... **CALL**
P341 SL .... **CALL**
P351 SX .... **CALL**

OKIDATA PRINTERS

320 .... **$350**
321 .... **499**
390 .... **499**
391 .... **699**
393 .... **999**
393 (color) **1099**
182 Plus/Parallel **239**

SWITCH BOXES

ROTARY

488024 2 pos. 25-pin AB **$15**
488058 2 pos. 36-pin AB **18**
448028 3 pos. 25-pin ABC **25**
448060 3 pos. 36-pin ABC **28**
448078 4 pos. 36-pin ABCD **30**
Other Switch Boxes .... **CALL**

PRINT BUFFERS

PRACTICAL PERIPHERALS

64K parallel or serial **$145**
256K parallel or serial **219**
64K parallel interface for Epson printers **75**
32K serial interface for Epson printers **75**

SMG

PB36 256 parallel **199**

MODEMS

HAYES and HAYES compatible

***HAYES Internal* (incl. software)**

1200B Baud **$279**
2400B Baud **429**

Compatible (including software)

Practical Peripherals 1200 baud **59.99**
Practical Peripherals 2400 baud **129**
Practical Peripherals 2400 MNP **175**
Hayes 1200 Baud **289**
Hayes 2400 Baud **439**
Everex 1200 baud **69**
Everex 2400 baud **139**

External (no software included)

U.S. Robotics 1200 baud **179**
U.S. Robotics 2400 baud **299**
U.S. Robotics 2400E **385**
U.S. Robotics 9600HST **599**
Prac. Peripherals 2400 baud Mini **79**
Practical Peripherals 2400 baud **179**
Practical Peripherals 2400MNP **199**

SMG smart money group **2400 Baud Modem**
- 100% Hayes Compatible - Internal - with ProComm Software $199 — **$99**
- 2 Year Warranty - External - no software $239 — **$129**

FOR TOSHIBA PORTABLES

2400 Baud Internal Modems

For T1000 .... **$209**
For 1100+, 1200, 3100, 5100 .... **199**

MONITORS

ZENITH

Zenith ZCM-1490 flat screen 14" VGA color monitor **$599**

SAMSUNG

Samsung Flat-screen Amber **115**

NEC

Multisync 2A color **499**
Multisync 3D **685**

AMDEK

410A, 410G, 410W 12" **150**
432 14" Mono Analog **159**
632 14" VGA **375**
732 12" Color **399**
735 14" Smartscan **499**

SONY

CPD 1303P Multiscan **435**

PACKARD BELL

12"Amber .... **$ 79**
14"EGA Color **335**
14"CGA Color **229**
14"Amber Flat Screen **99**
14"White Flat Screen **105**

DRIVES & BACKUPS

HARD DRIVES

SEAGATE

Internal Hard Disk Subsystems
20 Megabyte w/Contr. XT . **$275**
20 Megabyte 3.5" w/Cont. **319**
30 Megabyte RLL. XT **299**
40 Megabyte 40 MS ½HT **389**
40 Megabyte 28 MS ½HT **445**
80 Megabyte 28 MS **649**
123 Megabyte 28 MS with controller **899**

MINISCRIBE

30 Megabyte RLL XT **299**
40 Megabyte 28 MS ½HT **450**
70 Megabyte 28 MS ½HT **599**

HARD DRIVE CARDS

Western Digital 30 PS 30MB and 40MB .... **CALL**

PLUS DEVELOPMENT

20MB Hard Card **499**
40MB Hard Card **599**

TAPE BACKUPS

Mountain Internal

40MB XT or AT **379**
60MB AT **899**
60MB PS/2 .... **975**

Mountain External

40MB XT or AT **575**
150MB or AT PS/2 .... **1499**

Full Line of Mountain PC & PS/2 Tape Drives Available

DISK DRIVES

TOSHIBA

ND 04D ½ht. 360K DSDD (PC/XT) **$79**
ND 04DEG ½ht. 360K DSDD (PC/AT) **79**
ND 08DEG 1.2MB (PC/AT) **90**
ND 352KU 3.5" 720K (XT/AT) ... **90**
ND 356T 1.44MB (PC/AT) **115**

MOUSE DEVICES

Logitech

Serial Mouse **$68**
Hi-Res Bus II **86**

Microsoft

Paint brush Bus or Serial **98**
Easy CA Bus or Serial **127**
Paintbrush w/Windows Bus or Serial **135**

Mouse System

Mouse I & II

W/Autosketch Bus or Serial **107**
Pop-Up Menus Bus or Serial **87**
PS/2 .... **60**
Omni Mouse (Serial only) **34.99**
Mitsi "E-Mouse" **75**
Mouse Mat **3.99**

VIDEO CARDS

Paradise Short color RGB **$59**
Paradise Mono Graphics **79**
Paradise EGA 350 .... **159**
Paradise EGA 480 .... **179**
Paradise VGA Plus **249**
Paradise VGA Plus 16 .... **299**
Paradise VGA Professional **399**
Generic Hercules compatible **80**
Generic color compatible **79**
Hercules color **149**
Hercules Graphics PLUS **179**
Hercules inColor card **255***
Hercules VGA **189**
Vega Deluxe EGA **219**
Orchid Pro Designer VGA **325**

*$139 with purchase of any monitor.

POWER PROTECTION

TRIPPLITE & PERMA POWER

Surge Protector Strips

Isobar 4-6 4 outlet **$42**
Isobar 8-15 outlet **57**
Isobar 8-15RM 8 outlet w/remote **65**

Line conditioners/stabilizers

LS600 2 outlet 600 watt **85**
LC 1200 4 outlet 1200 watt **145**
LC1800 6 outlet 1800 watt **199**

Command consoles

Perma PD200 2 outlet **25**
Perma Power 6 outlet **28**
Perma RS562 Power Cmdr **65**
Perma RS610 .... **50**
CC1612 .... **69**
CC16-Plus built-in modem prot. **75**

Standby power systems

BC325 325 watt system **269**
BC450 450 watt system **299**
BC750 750 watt system **499**
BC1200 1200 watt system **650**
BC2000 2000 watt system **1199**

Adapter

PT209 .... **19.99**

ALL TRIPPLITE & PERMA POWER SUPPLIES IN STOCK AT DEEP DISCOUNTS

BOARDS & CHIPS

MULTIFUNCTION BOARDS

AST

Six Pak Plus 384 XT **$279**
Rampage 286 w/512K **399**
Rampage/2 PS/2 50&60 .... **549**
Advantage/2 PS/2 50&60 .... **499**
XFormer/286 .... **599**

ORCHID

Tiny Turbo 286 .... **235**
Ramquest PS/2 50/60 w/2MB **645**

INTEL ABOVEBOARD

Above Board Plus w/512K **399**
Above Board Plus I/0 w/512K **450**
Above Board 2 Plus for PS/2 models 50 & 60 .... **475**
Inboard 386 PC w/1 MB **599**
Inboard 386 PC IMB Piggyback **345**

MATH COPROCESSORS

Intel 8087 math (5 MHz) **$90**
Intel 80871 math (8 MHz) **175**
Intel 80872 math (6/8 MHz) **129**
Intel 80287 math (6/8 MHz) **139**
Intel 802878 math (8 MHz) **205**
Intel 8028710 math (10 MHz) **239**
Intel 8038716 math (16 MHz) **375**
Intel 8038720 math (20 MHz) **419**
Intel 8038725 math (25 MHz) **530**
Intel 80C287 for 80C286 .... **295**
NEC V-20-8 .... **9.99**

CHIPS

256K 80 Nanoseconds **$7.50**
256K 100 Nanoseconds **7.50**
256K 120 Nanoseconds **6.50**
256K 150 Nanoseconds **6.50**

DISKETTES

(*Packed in flip 'n' file) (**unformatted)			IBM			
3½" SSDD	–	12.00*	–	–	13.00	–
3½" DSDD	16.00	16.00	25.00	10.99	14.00	12.99
3½" DSDDHD	34.99	39.99	48.00	28.99	34.00	–
5¼" DSDD	8.00	8.00	15.00	5.99	–	5.99
5¼" DSDD96TPI	12.50	18.00	–	–	–	10.00
5¼" DSDDHD (for IBM AT)	12.50	14.00	30.00	9.99	–	14.00
8" SSDD**	19.00	20.00	–	–	–	–
8" DSDD**	24.00	23.00	–	–	–	–

3M

Diskettes

3½" SSDD	12.00
3½" DSDD	14.00
3½" DSDDHD	32.00
5¼" DSDD	8.00
5¼" DSDD96TPI	17.00
5¼" (for IBM AT) DSDDHD	14.00
8" SSDD**	20.00
8" DSDD**	24.00

Data Cartridges

DC100A	$13.50
DC300A	17.50
DC300XL P	21.50
DC600A	23.00
DC600HC	25.50
DC615HC	19.00
DC1000	14.50
DC1000 aloha 20MB*	15.50
DC1000 delta 20MB*	15.50
DC2000	17.00

*formatted cartridges
Quantity Discounts Available

FAX MACHINES

AT&T

3410D .... **$ 899**
3510D **1399**
3520D **1299**

SHARP

UX 180 .... **999**
UX 350 .... **$1399**

RICOH

RF800 .... **$679**

TOSHIBA

Model 3300 .... **$799**
Model 3700 .... **1199**

PERMA POWER STRIP FOR FAX

Perma RTD 410 .... **$45**

ELEK-TEK®, inc.

Corporate accounts invited, minimum order $15.00 Visa or MasterCard by Mail or Phone. Mail cashiers check, money order, personal check, add $4.00 1st item. (AK, HI, P.R., Canada add $10.00 1st item) $1.00 ea. add'l. shpg. and handl. Shipments to IL add 7% tax.

Write for free catalog DUNS #09-718-0517

IBM is a trademark of International Business Machines Co. Epson is a registered trademark of Seiko Epson Corporation.

6557 North Lincoln Avenue, Chicago Illinois 60645 Call in IL 312-982-5765
Call Toll Free 800-621-1269 — Except IL, AK Canadian Toll Free 800-459-9133

Prices subject to change, all products subject to availability.
ALL ELEK-TEK MERCHANDISE IS BRAND NEW, FIRST QUALITY AND COMPLETE.

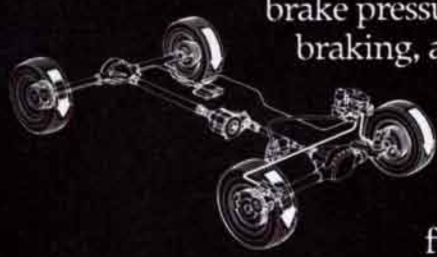
RETURN POLICY: All sales are final except defectives which will be replaced with identical merchandise only within 30 days. (computers and large peripherals — 3 work days) after which manufacturer's warranty applies.

© 1989 SCIENTIFIC AMERICAN, INC

This Year, Ou
Wanted to Stop Yo

Jeep announces our four-wheel anti-lock braking system.

It's called ABS. And it's a Jeep exclusive. It's the only all-wheel full-time braking system available on any four-wheel drive sport utility vehicle. ABS reads the road's surface condition up to fifteen times per second and automatically applies and reapplies brake pressure as needed. It allows drivers the freedom of steering while braking, and turns every stop, even panic stops, into more confident, more controllable stops. That's the beauty of ABS. But it's far from all the beauty of Cherokee Limited. There's the beauty of Selec-Trac, another Jeep exclusive. It's the only shift-on-the-fly four-wheel drive system that gives you the security of full-time four-wheel drive on any surface, on road or off. And,



© 1989 SCIENTIFIC AMERICAN, INC

r Designers
u in Your Tracks.

there's beauty under the hood in the form of the 177 horsepower Power-Tech Six engine, the most powerful engine in Cherokee's class.

Plus, there's beauty inside with luxurious seating highlighted by lots of genuine leather and surrounded by power everything. And maybe most important, there's the thrill and satisfaction of driving a true American Legend: Jeep. What's more, Jeep Cherokee has a higher resale value than Ford Bronco II and Chevy S-10 Blazer.* And

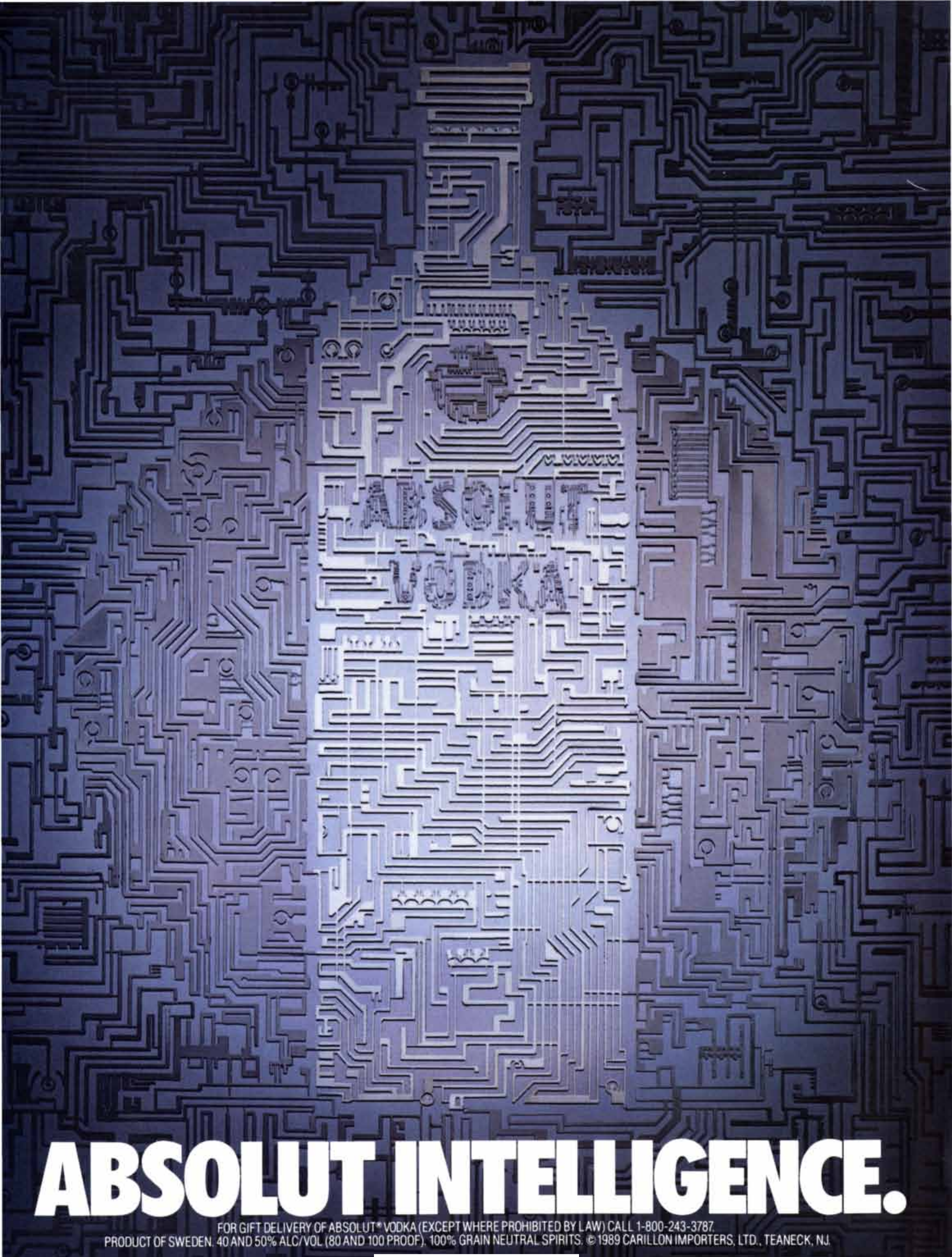
7/70 now every Jeep comes with Chrysler's exclusive 7-year/70,000-mile Protection Plan.** For further information, call 1-800-JEEP-EAGLE.

Only in a Jeep. Jeep Eagle

*Source: *Kelley Blue Book:* November/December 1988, analyzing resale values of 1985, '86, '87, and '88 model year vehicles.

**Protects 1989 models' engine and powertrain for 7 years or 70,000 miles and against outerbody rust-through for 7 years or 100,000 miles. See limited warranty at dealer. Deductibles and restrictions apply. Jeep is a registered trademark of Jeep Eagle Corporation. Buckle up for safety.

© 1989 SCIENTIFIC AMERICAN, INC



© 1989 SCIENTIFIC AMERICAN, INC